Strukturele modellen met theoretische variabelen

Verschuren, Petrus Johannus Maria

Publication date:
1980

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Verschuren, P. J. M. (1980). *Strukturele modellen met theoretische variabelen: de analyse van kausale relaties in modellen met theoretische variabelen door middel van stelsels lineaire strukturele vergelijkingen in het empirisch sociologisch en politikologisch onderzoek*. [, Tilburg University]. [s.n.].

STRUKTURELE MODELLEN MET THEORETISCHE VARIABELEN

P. VERSCHUREN

STELLINGEN

1. Een binnen de empirisch-analytische methode wijdverbreide werkwijze is het doen van uitspraken omtrent veranderingen binnen eenheden op basis van gekonstateerde verschillen tussen eenheden. De risiko's welke hieraan zijn verbonden, zijn niet in overeenstemming met de aandacht die wordt gegeven aan de methodologie en de toepassing van diachronisch onderzoek.

2. Indien het kreëren van een nieuwe werkelijkheid datgene is wat wetenschap gemeenschappelijk heeft met kunst, dan kan hieruit worden geleerd dat het wetenschappelijk onderwijs vooral dient te zijn gekenmerkt door een aktief bezig zijn met materiaal, eerder dan door kennisoverdracht sec. Deze stelling is zelfs van kracht wanneer men onder het kreëren van een nieuwe werkelijkheid niet meer wenst te begrijpen dan het formuleren van wetenschappelijke theorieën.

3. Veel resultaten van wetenschappelijk onderzoek kunnen in een ander daglicht komen   staan wanneer men zich bij het bepalen van de maatschappelijke relevantie ervan niet uitsluitend baseert op een beoordeling van de mate waarin onderzoeksresultaten toepasbaar zijn op maatschappelijke problemen, maar tevens op de verwachting van de mate waarin deze resultaten daadwerkelijk een uitwerking zullen hebben.

4. Het mag worden verwacht dat feministisch onderzoek op basis van materiaal dat is verzameld op één bepaald tijdstip, zal resulteren in een uitbreiding van feministische theorieën alsook in een verhoogde maatschappelijke relevantie; het laatstgenoemde met name indien dit tijdstip wordt gekozen anno 1980.

5. Gevreesd moet worden dat de nadruk welke de minister van onderwijs en wetenschappen voornemens is te gaan leggen op de samenwerking tussen wetenschappelijk onderzoekers zoals is uitgewerkt in de BUOZ-nota, zal leiden tot het plaatsen van onderzoeksprojekten onder abstrakte gemeenschappelijke noemers, eerder dan tot het schrijven van gemeenschappelijke publikaties.

6. Als het waar is dat tegenstellingen elkaar aantrekken en gelijkenissen elkaar afstoten, dan ligt hierin wellicht de verklaring voor het feit dat in onderwijsprogramma's van sociale wetenschappen de vakken statistiek en methoden van onderzoek nog steeds niet zijn geïntegreerd.

7. In een tijd dat in Nederland de P.K. wordt afgeschaft als eenheid van energie, getuigt de uitbreiding van het aantal paarden in dit land van een grote luxe.

8. Het veelvuldig voorkomen van promotiefeesten is een van de weinige voordelen van het op individuele wijze schrijven van proefschriften.

STRUKTURELE MODELLEN MET THEORETISCHE VARIABELEN

De analyse van kausale relaties in modellen met theoretische variabelen door middel van stelsels lineaire strukturele vergelijkingen in het empirisch sociologisch en politikologisch onderzoek.

Promotores:

Prof. dr. Ph.C. Stouthard

Prof. dr. A.L. Hempenius

Prof.dr. R.J. Mokken heeft, met name in de beginfase, als begeleider bijgedragen aan de totstandkoming van deze studie.

STRUKTURELE MODELLEN MET THEORETISCHE VARIABELEN

De analyse van kausale relaties in modellen met theoretische variabelen door middel van stelsels lineaire strukturele vergelijkingen in het empirisch sociologisch en politikologisch onderzoek.

Proefschrift ter verkrijging van de graad van doctor in de sociale wetenschappen aan de Katholieke Hogeschool Tilburg, op gezag van de Rector Magnificus, Prof. Dr. J.E.A.M. van Dijck, in het openbaar te verdedigen ten overstaan van een door het College van Decanen aangewezen commissie in de aula van de Hogeschool op donderdag 5 juni 1980 te 16.00 uur

door

Petrus Johannus Maria Verschuren

Geboren te Heusden (N.Br.)

methode = weg waarlangs

INHOUD

Het is Anneli Vos geweest die het lastige typewerk met uitzonderlijke vaardigheid heeft verricht.

INLEIDING

Een klacht die men veel hoort bij beoefenaars van sociale wetenschappen betreft het bestaan van een kloof tussen theorie en onderzoek. Met enig cynisme zou men kunnen stellen dat de scheiding van arbeidsterreinen slechts een andere formulering is van hetzelfde probleem. Zij brengt ons evenmin dichter bij een oplossing gezien de niet te stuiten specialisering van wetenschappelijke arbeid. Een nadere bezinning op de praktijk van zowel theorievorming als onderzoek lijkt een meer vruchtbare weg. Bekijken we eerst het probleem vanuit de theorie. Het grootste deel van wat in sociale wetenschappen als politikologie en sociologie heeft plaatsgevonden, is de formulering van verbale theorieën. Deze bevatten meestal een verzameling nominaal gedefinieerde begrippen alsmede een verzameling uitspraken waarin relaties worden gelegd tussen de begrippen.

Allereerst de relaties. In veel theorieën geven deze aanleiding tot komplikaties. Hoewel soms met een relatie niets anders is bedoeld dan een meer dan toevallig samengaan van verschijnselen, gaat de betekenis van het relatiebegrip in veel gevallen verder. Zo spreken theoretici over "vloeit voort uit", "is geïnduceerd door", "is een effekt van" en dergelijke. Ondanks het feit dat vele, met name positivistische, wetenschappers anders willen doen geloven, meen ik dat hier onontkoombaar sprake is van een kausale denkwijze. Reden dat zo weinig theoretici het begrip kausaliteit expliciet hanteren, zou kunnen zijn het feit dat het een uitermate lastig begrip is. Om dezelfde reden geven zij ook zelden precies aan wat wordt verstaan onder kausaliteit, noch behandelen zij de vraag wat de betekenis is van het begrip oorzaak in en voor hun theorieën.

Ook de tweede komponent van theorieën, de begrippen, zorgen vaak dat de afstemming van theorie en onderzoek problematisch is. Zo zijn veel theorieën ingewikkeld als gevolg van het grote aantal gehanteerde begrippen. Dit leidt tot een dermate komplex relatiepatroon dat een adekwate empirische toetsing nauwelijks uitvoerbaar is. Merton lanceerde in dit kader zijn idee van "middle-range" theorieen. Een ander minstens zo belangrijk probleem is dat begrippen veelal vaag en van een zeer hoog abstraktienivo zijn. Dit is temeer een probleem daar in de theorieën zelf slechts bij uitzondering aanwijzingen zijn te vinden voor de operationalisering van begrippen. Slaagt de onderzoeker niet in een adekwate operationalisering, dan is een steekhoudende empirische toetsing van de theorie niet mogelijk. Worden immers de verwachte relaties niet in de empirie gevonden dan blijft onzeker of dit komt doordat de theoretische relaties niet kloppen of doordat de vertaling van theoretische begrippen in onderzoeksvariabelen op invalide of onbetrouwbare wijze is geschied.

Bezien we vervolgens het probleem vanuit de onderzoekspraktijk. Behoudens allerlei ad hoc procedures voor operationalisering, meting, logische en empirische toetsing zijn voor de ontwikkeling van politikologische en sociologische theorieen vooral van belang multivariate benaderingen als simulatie-modellen, graphen-theorie en kausale modellen. Gezien een eerdere konstatering aangaande de aard van theoretische relaties, lijken met name kausale modellen geschikt voor ondersteuning van theoriekonstruktie.

Het doel van deze studie is een methodologische bestudering van de toepassing van kausale modellen als representatie van theorieën. Meer in het bijzonder gaat het om schattingsproblemen welke optreden zodra theoretische variabelen worden opgenomen in een kausaal model.

Alvorens hiertoe over te gaan, komen in de eerste twee hoofdstukken problemen aan de orde welke indirekt aan de vermelde problematiek zijn verbonden. Voor kausale modellen geldt in het algemeen een onvoldoende explicitering van het kausaliteits- en modelbegrip. In hoofdstuk 1 wordt getracht enige duidelijkheid dienaangaande te scheppen.

Een tweede probleem betreft de in het hedendaags politikologisch en sociologisch onderzoek meest toegepaste techniek voor analyse van kausale modellen, te weten de pad-analyse. Het is onvoldoende bekend dat deze techniek kan worden gezien als een bijzonder geval van een in de ekonometrie reeds ver uitgewerkte theorie, te weten die van strukturele vergelijkingen. Dit wordt in hoofdstuk 2 verduidelijkt, terwijl in hoofdstuk 3 de analyse van modellen met theoretische variabelen wordt behandeld, waarbij psychometrische en ekonometrische inzichten en procedures op kritische wijze aan de orde worden gesteld.

Een fundamentele kritiek op procedures zoals die zijn uitgewerkt in de psychometrie, ekonometrie en sociologie, procedures bedoeld voor analyse van modellen met theoretische variabelen, volgt in hoofdstuk 4. De lijn waarlangs deze kritiek zich beweegt, wordt gevormd door een nadere analyse van het begrip meetfout. In de traditionele behandeling van de meetfoutenproblematiek wordt deze geacht geheel en al te zijn gelokaliseerd in de fase van dataverzameling en -verwerking. Dit heeft tot gevolg dat de aandacht vrijwel uitsluitend is gericht op de kategorie van random meetfouten. Het begrip meetfout kan evenwel worden uitgebreid in die zin dat elke verstoring in de representatie van theoretische begrippen door onderzoeksvariabelen als een meetfout kan worden opgevat. Dat betekent dat deze verstoring zich niet alleen in beide genoemde fasen kan voordoen, maar tevens tijdens de fasen van operationalisering en analyse. Deze verstoring heeft steeds een systematisch karakter, zodat kan worden gesproken van *nonrandom* meetfouten. Dit houdt in een uitbreiding van meetfouten als betrouwbaarheidsprobleem naar een geldigheidsprobleem.

Dergelijke meetfouten treden onder andere op zodra de analyse van een kausaal model met meervoudig geoperationaliseerde theoretische variabelen wordt uitgevoerd op basis van afzonderlijke indikatoren. Zoals analyses in hoofdstuk 4 laten zien, ontstaat aldus een mengeling van strukturele en epistemische aspekten, welke de semantische eigenschappen van theoretische variabelen ongunstig beïnvloeden. In de hoofdstukken 5 en 6 wordt als oplossing voor dit probleem voorgesteld de analyse van konseptmodellen voorafgaand aan de analyse van een struktureel model als geheel.
Bij de bespreking van het kausaliteitsbegrip in hoofdstuk 1 heb ik niet de pretentie van filosofische deskundigheid. Veeleer is het de bedoeling na te gaan hoe het kausaliteitsbegrip met het oog op sociologisch en politikologisch empirisch onderzoek op zinvolle wijze kan worden omschreven.

De statistische problemen betreffende de relatie steekproef-populatie blijven goeddeels buiten beschouwing. De diskussies worden gevoerd op modelnivo waardoor een notatie in parameters in plaats van steekproef-grootheden is gerechtvaardigd. Slechts in gevallen waarin een bespreking en notatie op het nivo van waarnemingen onvermijdelijk lijkt, is zulks gedaan.

Veel aandacht is besteed aan het ontwikkelen van een begrippenkader. Zo wordt gesproken over hoofd- en hulptheorie, strukturele en epistemische relaties, strukturele en epistemische bias, errorvrije en latente variabelen, een typologie van meetfouten en konseptmodellen. De achterliggende gedachte is dat de bruikbaarheid van multivariate technieken in sterke mate afhangt van de aanwezigheid van een bijpassend begrippenapparaat.

Tenslotte de volgende opmerking: elke procedure voor het onderzoeken van kausale hypothesen, inklusief de door mij uitgewerkte, brengt risiko's mee. Dit bij gebrek aan methoden met behulp waarvan een sluitende bewijsvoering mogelijk is. Dit zal in nog sterkere mate gelden voor modellen waarin ongemeten variabelen zijn opgenomen. Deze konstatering betekent evenwel allerminst een aantasting van het bestaansrecht van deze studie. Wél geeft zij aanleiding tot het besef dat het maximaal haalbare van kausaal onderzoek steeds is de aanname van kausale hypothesen zo plausibel mogelijk te maken. De risiko's welke hierbij optreden zijn even noodzakelijk als onvermijdelijk.

1. KAUSALITEIT EN KAUSALE RELATIES

In de afgelopen decennia is er bij wetenschappers een tendens merkbaar naar het uitbannen van het begrip kausaliteit. Liever spreekt men over funktionele relaties, voorspelling en wetmatigheid dan zich te wagen aan de voetangels van kausaliteit. Het meest duidelijk heeft deze tendens zich voorgedaan in de fysika en, naar aanleiding van ontwikkelingen hierin, tevens in de natuurfilosofie. Hier ontwikkelde zich gedurende de 17e eeuw een kausaliteitsbegrip dat sterk verwant is met de begrippen wetmatigheid en voorspelbaarheid. Deze opvatting, met name aangehangen door Laplace, wordt aangeduid met de term "mechanicisme", dit vanwege het feit dat de gebeurtenissen in de natuur werden gezien als de opeenvolgende fasen in een mechanisch proces. Zij volgen elkaar in onwrikbare volgorde op. Geheel passend in de klassieke visie op het natuurgebeuren is het postulaat van absolute voorspelbaarheid. In het mechanicistische wereldbeeld van Laplace zou iemand die de momentele toestand van het universum precies zou kennen, en bovendien alle wetten die in dit universum gelden ter beschikking had, elke gebeurtenis in het verleden en in de toekomst exakt kunnen berekenen, dan wel voorspellen.

Tegen deze veronderstelling van absolute voorspelbaarheid kwamen bij de opkomst van de kwantummechanika aan het begin van de twintigste eeuw steeds meer bezwaren. De bezwaren hielden in dat er in de realiteit sprake is van een absoluut indeterminisme ofwel van een primaat van het toeval. Volgens de kwantumtheorie is het onmogelijk om tegelijkertijd de positie en de snelheid van een subatomair deeltje exakt te bepalen. De exakte bepaling van de positie sluit de bepaling van de snelheid uit en omgekeerd. In de kwantummechanika noemt men dit het beginsel van onbepaaldheid, ook wel bekend als de "onzekerheidsrelaties van Heisenberg".

Uit dit beginsel nu hebben velen gemeend te kunnen afleiden een principiële onvoorspelbaarheid van de gebeurtenissen en als gevolg daarvan een afwijzing van het kausaliteitsbeginsel. Indien de gebeurtenissen elkaar op absoluut onvoorspelbare wijze opvolgen, is het onmogelijk dat deze gebeurtenissen gehoorzamen aan kausale wetten. Deze konklusie welke door fysici werd getrokken en later werd overgenomen door neopositivisten als Carnap, Russell en Schlick, wordt door zeer velen bestreden. Men zie bijvoorbeeld Van Melsen (1961), Bunge (1959), Welten (1961) en Nagel (1961). Hun tegenargumenten betreffen voornamelijk de onjuistheid van gelijkstelling van kausaliteit en voorspelbaarheid, het zelfkontradiktorische van ordeloosheid in de natuur en de stelling dat determinisme niet onverenigbaar is met het statistische karakter van wetten in de kwantummechanika.

Mede beïnvloed door genoemde ontwikkelingen in de fysika, ontwikkelden neopositivisten een kausaliteitsbegrip dat praktisch samenvalt met wetmatigheid en voorspelbaarheid. Van kausaliteit is volgens hen slechts sprake daar waar door middel van wetten de toekomstige ontwikkeling van gebeurtenissen kan worden voorspeld op basis van de huidige toestand en de thans geldende wetten. Zo stelt Carnap (1969: 193): "Ein Ereignis Y zur Zeit T wird von einem vorhergehenden Ereignis X dann und nur dann verursacht wenn Y aus X mit Hilfe der Gesetz G, die zur Zeit bekannt sind, ableitbar ist." Het is deze erfenis uit de natuurwetenschappen die waarschijnlijk verantwoordelijk is voor een tendens in de sociale wetenschappen naar uitbanning van het kausaliteitsbeginsel.

1. KAUSALE RELATIES

De toonzetting hierboven doet reeds vermoeden dat de zojuist beschreven opvatting in dit geschrift niet wordt gedeeld. Het meest wezenlijke van kausaliteit lijkt mij een *produktiekarakter* te zijn. Van een kausale relatie is in deze opvatting eerst sprake wanneer de oorzaak het gevolg produceert. Als X de oorzaak is van Y dan wordt Y door X tot stand gebracht. Deze opvatting is evenmin nieuw als zonder problemen. Het grootste probleem is gelegen in de principiële onmogelijkheid tot waarneming van het proces van produceren. Wat wordt waargenomen is een bij herhaling optredende tijd-ruimtelijke nabijheid van verschijnselen zoals verhitten-uitzetten, zaaien-oogsten, ziekte-dood etc. Het feitelijke produceren, het proces waarbij de oorzaak het gevolg tot stand brengt, onttrekt zich aan de waarneming. Dit is voor met name neopositivisten mede aanleiding geweest het begrip oorzaak te schrappen uit het wetenschappelijk vokabulaire en te vervangen door begrippen als wetmatigheid en voorspelbaarheid. Echter veeleer dan de realiteitszin van het begrip kausaliteit in twijfel te trekken, ligt het voor de hand te zoeken naar methoden waarmee een produktiekarakter van een relatie tussen verschijnselen meer plausibel kan worden gemaakt.

Voor een nadere toespitsing van de wijze waarop het begrip kausaliteit inhoud wordt gegeven binnen empirische wetenschappen in het algemeen en het empirisch politikologisch en sociologisch onderzoek in het bijzonder, is het nodig in te gaan op een specifieke taal welke daar gebezigd wordt, te weten een variabelentaal. Kenmerkend hiervoor is dat het objekt van onderzoek wordt gedefinieerd als een verzameling eenheden welke worden gezien als dragers van eigenschappen. Bijvoorbeeld in de fysika kan men ijzer zien als een eenheid met als eigenschappen (ofwel toestandkenmerken) temperatuur, hardheid, soortelijk gewicht, uitzettingskoëfficiënt etc. In de politikologie neemt men individuen of gehele maatschappijen als eenheden met bijvoorbeeld religie, demokratische gezindheid respektievelijk politieke struktuur en welvaartsnivo als kenmerken. Een variabele nu is een kenmerk dat verschillende vormen (nominale variabele) dan wel gradaties (ordinale en metrische variabele) kan aannemen.

Er zijn twee manieren waarop verschillende vormen dan wel gradaties van een kenmerk zich presenteren: (a) verschillende eenheden hebben verschillende vormen of gradaties, (b) bij een en dezelfde eenheid treedt door de tijd heen een verschuiving op in vormen of gradaties. Uit het bovenstaande volgt een tweede element uit de empirische onderzoekspraktijk dat van belang is voor een adekwate omschrijving van het begrip kausaliteit, te weten dat van *verandering*. Meestal is men geïnteresseerd in de vraag of een verandering in de ene eigenschap steeds samengaat met een verandering in de andere. Zo mogelijk tracht men te achterhalen of de ene verandering de andere tot stand brengt. Bijvoorbeeld, welke zijn de te verwachten veranderingen in arbeidsbevrediging als gevolg van verbetering van een inspraakstruktuur? Aldus ontstaat een kausaliteitsbegrip waarbij het niet zozeer gaat om relaties tussen verschijnselen maar tussen *veranderingen* daarin.[1)]

Een kausale relatie zoals hierboven gedefinieerd, heeft sterke raakvlakken met het mathematische begrip funktie alsook met een relatie zoals wordt weergegeven in een regressievergelijking. Zij leidt tot de konstatering van de volgende serie kenmerken. Allereerst de faktor *tijd*. Met name sociale processen verlopen met enige traagheid, waardoor vrijwel steeds een tijdsverschil aanwezig is tussen het optreden van oorzaak en gevolg. Een tweede reden waarom het gevolg zich eerst enige tijd na de oorzaak voordoet, heeft betrekking op indirekte oorzaken. Zo is X een indirekte oorzaak van Z indien X een invloed heeft op Y en Y weer op Z. Hierover aanstonds meer bij de bespreking van transitiviteit.

Een tweede kenmerk van kausale relaties is een *kansaspekt*. Meestal wordt de vraag of kausaliteit deterministisch dan wel probabilistisch is, beantwoord vanuit een veronderstelde wetmatigheid van verschijnselen (neopositivisme). Een poging deze vraag te beantwoorden vanuit een meer ontologische opvatting van kausaliteit, leidt tot de stellingname dat een kanskarakter van kausaliteit kan worden afgeleid uit het feit dat kausale processen zich in de tijd voltrekken. Immers na aanvang van de werking van oorzaak X kunnen additionele oorzaken in werking treden die het kausale proces een andere wending doen nemen. Het gevolg is dat Y hetzij niet optreedt hetzij er anders uit komt te zien dan zonder tussenkomst van deze additionele oorzaken het geval zou zijn geweest. Voor zover deze additionele oorzaken niet samenhangen met de initiële oorzaak, dat wil zeggen noch een kausale samenhang noch een schijnverband hebben met X, zorgen zij voor een stochastisch element in de kausale relatie.

Het meest overtuigende argument voor een probabilistische opvatting van kausaliteit is de aanname van een menselijke vrije wil (zie van Melsen: 1946). Een niet door de mens gewenst gevolg van een oorzaak kan hij soms verhinderen op te treden. Dit niet uitsluitend door te voorkomen dat een oorzaak zich voordoet. Dankzij het tijdsaspekt dat verbonden is aan de meeste kausale processen, is het soms mogelijk een oorzaak, nadat die zich voordeed, ongedaan te maken. Er lijken dan ook voldoende redenen aanwezig om in het empirisch kausaal onderzoek niet te werken met een deterministische opvatting van kausaliteit maar met een probabilistische. Dat wil zeggen, wanneer een verandering in oorzaak X optreedt dan wordt aangenomen dat met een zekere mate van waarschijnlijkheid gevolg Y onder invloed van X een verandering ondergaat. Of, meer kwalitatief geformuleerd, wanneer X zich in hoedanigheid x voordoet dan zal onder overigens gelijkblijvende omstandigheden Y zich in een groter aantal gevallen in hoedanigheid y voordoen dan wanneer X zich niet als zodanig voordeed.

Een aspekt van kausaliteit, dat wel eens wordt verward met het kans-aspekt, is de *sterkte* van kausale werkingen. Het betreft hier evenwel twee volkomen onderscheiden begrippen. De sterkte van een kausale relatie houdt in de hoeveelheid verandering in een kausaal afhankelijke variabele welke optreedt als gevolg van bepaalde veranderingen in de oorzaak. Wanneer X een oorzaak is van Y, en Y verandert slechts een beetje onder invloed van X, dan is er sprake van een zwakke kausale relatie. Maar daarom is deze relatie nog niet probabilistisch en kan zij best in honderd percent van de gevallen opgaan.

Hoewel veel auteurs stellen dat kausaliteit een zaak is van alles of niets, zijn er verschillende argumenten voor de aanname van een variabele sterkte van kausale afhankelijkheden. Zo kan er sprake zijn van verschillende oorzaken. Dit heeft op twee manieren konsekwenties voor de sterkte van kausale afhankelijkheid. Allereerst is het mogelijk dat de sterkte toeneemt wanneer meerdere oorzaken uit de totale verzameling oorzaken zich tegelijkertijd voordoen. Zelfs zal in het geval de oorzaken interakteren en zij een konjunktieve set vormen, de sterkte van hun gezamelijke werking op multiplikatieve wijze toenemen. Ten tweede is het mogelijk dat oorzaken elkaar geheel of gedeeltelijk neutraliseren. Wanneer bijvoorbeeld zowel een stoffige als een vochtige werkruimte oorzaken zijn van arbeidsontevredenheid, dan kunnen deze oorzaken elkaar (gedeeltelijk) ongedaan maken.

Vervolgens is er nog de veelomstreden kwestie van symmetrie versus asymmetrie. De meeste auteurs zijn van oordeel dat een kausaal proces asymmetrisch is. Dit lijkt overduidelijk wanneer we gaan zien naar de wijze waarop kausaliteit op verreweg de meest betrouwbare wijze kan worden onderzocht, te weten via een manipulatief handelen. Als bijvoorbeeld blijkt dat na een wijziging van X een verandering optreedt in Y, dan is de verandering in X de oorzaak van de verandering in Y en niet omgekeerd. Niettemin komt men in de literatuur wel aanhangers tegen van het standpunt dat sommige kausale relaties symmetrisch zijn. Zo is het in hun redenering best mogelijk dat tussen verschijnselen als inkomen en opleiding in die zin een symmetrische relatie bestaat, dat door een verhoging van opleiding een toename van iemands inkomen kan worden bewerkstelligd. Omgekeerd is het denkbaar dat iemand meer opleiding gaat volgen als gevolg van een stijging van het inkomen. Dit laatste bijvoorbeeld vanwege verruiming van voor opleiding benodigde financiële middelen of vanwege een behoefte aan konformatie aan mogelijk bij hogere inkomensgroepen bestaande waarden en normen op dit punt.

Er zijn een tweetal redenen die tegen deze opvatting van symmetrie pleiten. Allereerst is hier ongemerkt een variabelentaal binnengeslopen. Dat wil zeggen, er werd een uitspraak gedaan omtrent de relatie tussen inkomen en opleiding *opgevat als klassen van verschijnselen*. Het valt echter niet in te zien hoe tussen klassen van verschijnselen een kausale relatie kan bestaan, aangezien deze klassen abstrakties zijn. Kausaliteit speelt zich echter af in konkrete processen, hetgeen in dit geval wil zeggen dat er voor een *konkreet* persoon een kausale relatie kan bestaan tussen diens inkomen en opleiding. De bewering dat tussen de variabelen inkomen en opleiding een kausale relatie bestaat is dan ook in feite zinledig. Het is op het nivo van afzonderlijke eenheden, als elementen van de verzamelingen waarop variabelen zijn gedefinieerd, dat kausale processen zich voltrekken.

Een tweede tegenwerping tegen de aanname van symmetrie is dat een kausaal proces in de ene richting waarschijnlijk van een ander karakter is dan het proces dat zich in tegenovergestelde richting voltrekt. Bijvoorbeeld, het kausale proces volgens hetwelk toename van inkomen leidt tot een toename in opleiding is van een wezenlijk ander karakter dan het proces van beïnvloeding dat mogelijkerwijs uitgaat van toename in opleiding naar toename in inkomen. In beide gevallen is sprake van verschillende procesvariabelen, hetgeen betekent dat de kausale mechanismen geheel verschillend zijn van aard.

Een vraag die zich voorts opdringt is die naar de positieve respektievelijk negatieve werking van een oorzaak. Heeft X een toename of juist een afname van Y tot gevolg? Het betreft hier de *valentie* als kenmerk van kausale relaties. Deze is uiteraard slechts van toepassing op relaties tussen niet-nominale variabelen.

Voor de hantering van technieken uit de hoofdstukken hierna is het tot slot noodzakelijk uit te gaan van een *transitieve* eigenschap van kausale relaties. Deze is van belang in het geval van indirekte oorzaken. Stel dat een indirekte relatie tussen X en Z tot stand komt via Y. Symbolisch

$$X \longrightarrow Y \longrightarrow Z$$

Deze relatie is opgebouwd uit twee direkte relaties namelijk die tussen X en Y en die tussen Y en Z. De eigenschap van transitiviteit nu houdt in dat uit beide direkte relaties voortvloeit dat X een oorzaak is van Z. De reden waarom transitiviteit thuishoort in een omschrijving van het begrip kausaliteit zoals in deze studie gehanteerd, is dat in de hoofdstukken hierna een relatie als tussen X en Z hierboven een modelparameter krijgt toegewezen, de zogenaamde gereduceerde-vorm-parameter. Deze wordt normaliter kausaal geïnterpreteerd hetgeen uitsluitend mogelijk is onder de veronderstelling van transitiviteit.

Bovenstaande diskussie betreft in wezen het bestaan van kausale ketens. Hierbij twee opmerkingen welke beide van belang zijn voor een juist begrip van de hier gebezigde kausale denkwijze. Allereerst lijdt het geen twijfel dat gebeurtenissen steeds deel uitmaken van langere kausale ketens. Een oorzaak heeft meestal zelf oorzaken en een gevolg heeft zelf gevolgen. Ten tweede moet worden gesteld dat een direkte kausale relatie bij nader onderzoek vaak een indirekte blijkt te zijn. Ook Welten benadrukt het bestaan van tussenschakels in het proces van veroorzaking waar hij zegt: "hoewel oorzaak en gevolg niet gelijktijdig kunnen zijn, moeten zij wel op elkaar aansluiten, direkt of indirekt", en voegt daar als konklusie aan toe: "dit maakt een kontinue analyse van het tussengebeuren noodzakelijk".

Dit laatste wordt hier onderschreven. In paragraaf 3 wordt de opvatting uiteengezet dat het uiteenleggen van korrelaties in kausale ketens moet worden gezien als een belangrijke vorm van kausale verklaring.

In het bovenstaande heb ik getracht een kausaliteitsbegrip te omschrijven dat goede perspektieven biedt voor het empirisch sociaal-wetenschappelijk onderzoek. Het resulterende kausaliteitsbegrip is nauw gerelateerd aan een variabelentaal, waardoor een verbinding ontstaat met het begrip funktie uit de wiskunde en het relatiebegrip uit de regressietheorie. Een probleem in de gegeven begripsomschrijving wordt evenwel gevormd door het *veranderingsaspekt* alsook het *produktiekarakter* , temeer daar in deze studie uitsluitend gesproken wordt over de analyse van data verzameld op één tijdstip. Beide bemoeilijken het op adekwate wijze verkrijgen van kennis omtrent kausale relaties.

2. KENNIS VAN KAUSALE RELATIES

Bij de vraag hoe op empirische wijze kennis kan worden verkregen omtrent kausale relaties, is het goed na te gaan wat diegenen die zich in hun wetenschapsbeoefening sterk baseerden op de waarneming hierover te zeggen hebben. Hume is wel diegene geweest van de empiristen die verreweg de meeste invloed op het huidige kausaliteitsdenken heeft gehad. Hij bracht kausaliteit terug tot de konstante opeenvolging van verschijnselen. Meer precies is het kausaliteitsbeginsel bij hem gebaseerd op het geassocieerd raken van voorstellingen van waarneembare verschijnselen, wanneer bij herhaling wordt waargenomen dat het ene verschijnsel onmiddellijk door een ander wordt gevolgd. Naast konstante opeenvolging onderscheidt hij nog als tweede kriterium de tijdruimtelijke nabijheid van twee verschijnselen. Dat konstante opeenvolging en tijdruimtelijke nabijheid geschikte kriteria zijn voor de konstatering van kausaliteit zoals door Hume opgevat, blijkt hieruit dat zij inderdaad voorwaarden zijn voor de associatie van voorstellingen in de geest. Niet voor niets zijn dit de voornaamste grondpijlers geweest van de later door psychologen uitgewerkte theorie van klassieke konditionering.

Nadere beschouwing van het kriterium van konstante opeenvolging laat zien dat het hier in feite gaat om twee kriteria: (a) konstant samengaan en (b) tijdsvolgorde. Is de eis van tijdsvolgorde, welke behelst dat het gevolg per se moet komen ná de oorzaak, zonder meer duidelijk, die van konstant samengaan is dat zeker niet. In sommige gevallen zullen twee verschijnselen kausaal gerelateerd zijn zonder dat beide steeds samengaan. Zo zal slechts in uitzonderingsgevallen na bliksem brand ontstaan. Konstant samengaan lijkt dan ook bezwaarlijk te kunnen worden gezien als een noodzakelijke voorwaarde voor het konstateren van kausaliteit. Evenmin is het een voldoende voorwaarde. Bijvoorbeeld, twee machines die op puur toevallige wijze steeds vlak na elkaar een bepaalde beweging uitvoeren. Zelfs behoeft de konstante opeenvolging niet toevallig te zijn om ook niet kausaal te zijn. Bijvoorbeeld, twee klokken die steeds vlak na elkaar slaan. Deze regelmaat, hoewel niet toevallig aangezien beide systemen gericht zijn op een objektieve tijdregistratie, is niet kausaal bepaald.

Hoewel konstant samengaan als een noodzakelijke noch voldoende voorwaarde voor de konstatering van kausaliteit lijkt te kunnen worden beschouwd, is dit niettemin vaak een indikatie voor het bestaan van een kausale relatie. Zo zullen weinigen aarzelen om konstant samengaan zoals tussen rook en vuur, klap en boosheid, ouderdom en fysieke aftakeling, kausaal te interpreteren. Het is niet goed te zeggen waarom. Wellicht is het onze verwachting met vuur rook en met een klap boosheid te kunnen produceren die hier de doorslag geeft.

Onmiddellijk rijst de vraag hoe enige zekerheid te verkrijgen is omtrent het bestaan van een *produktiekarakter*. Een tweede probleem dat ten aanzien van konstant samengaan dient te worden besproken is het aspekt van *verandering* als onderdeel van de eerder gegeven kausaliteitsomschrijving. Het middel bij uitstek om wat meer zekerheid omtrent de aanwezigheid van een produktiekarakter te verkrijgen is het experiment. Hier is steeds de onderzoeker ultieme oorzaak als toediener van stimuli (oorzaken), waardoor een sterke subjektieve kausaliteitservaring kan optreden. Deze berust in niet onbelangrijke mate op tijdruimtelijke nabijheid alsook op het principe van konstante opeenvolging. Dat in het experiment het element van konstante opeenvolging impliciet aanwezig is, behoeft enige toelichting. Vatten we het toedienen van de stimulus alsook het optreden van de response op als dichotome variabelen, dan kan men zeggen dat in het experiment wordt nagegaan in hoeverre er sprake is van een konstant samengaan van deze beide variabelen.

Behalve de subjektieve kausaliteitservaring geeft ook aselektief toedelen een extra mogelijkheid voor het meer plausibel maken van een produktiekarakter. Mogelijke storende invloeden welke de konklusie zouden kunnen verstoren, worden aselektief toegedeeld.

Naast genoemde voordelen zijn de nadelen van het experiment eveneens genoegzaam bekend. Zo zijn er de kunstmatigheid, de geringe generaliseerbaarheid, de moeilijke praktische uitvoerbaarheid en de onmogelijkheid bepaalde storende effekten te randomiseren. Genoemde bezwaren spelen vooral de sociale wetenschappen parten, om welke reden hier veelvuldig is gezocht naar niet-experimentele vormen van kausaal onderzoek. In de sociologie en politikologie heeft dit geresulteerd in procedures waarbij het begrip konstant samengaan, operationeel gemaakt via associatie-maten, veel meer expliciet dan het geval is in het experiment tot uitgangspunt dient. Bedoeld zijn elaboratie-, Simon-Blalock- en pad-analyse.

De vraag is of en in hoeverre hier het aspekt van verandering als onderdeel van de eerder gegeven kausaliteitsomschrijving recht wordt gedaan. Bij het toetsen van de hypothese - X is een oorzaak van Y - zou men als volgt te werk kunnen gaan. Stel Δ_x en Δ_y geven aan de verandering welke in X respektievelijk Y optreedt binnen een bepaald tijdsverloop. De vraag of Δ_x en Δ_y samenhangen als onderdeel van de kausaliteitsvraag, kan worden beantwoord via berekening van de korrelatie $r(\Delta_x \Delta_y)$. Is deze signifikant afwijkend van nul dan wordt in afwachting van verdere analyses de kausale hypothese voorlopig bevestigd geacht.

Bij de berekening van korrelaties op basis van "*cross-sectional*" data gebeurt wel iets dergelijks, maar het element van verandering heeft hier een iets andere betekenis. Om dit te laten zien wordt eerst de korrelatie-koëfficiënt op een van het normale gebruik afwijkende wijze geschreven. Stel de afzonderlijke eenheden i (i=1,...,N) in de onderzoekspopulatie hebben skore x_i respektievelijk y_i (i=1,...,N) op de beide variabelen. Stel voorts $\Delta_x = x_i - x_j$ en $\Delta_y = y_i - y_j$ (i,j=1,....,N; i≠j) zijn de verschilskores. Let wel, het gaat nu niet om verschillen *binnen* , maar *tussen* eenheden! Het kan gemakkelijk worden bewezen dat de traditionele korrelatie-koëfficiënt r(XY) nu als volgt kan worden geschreven

$$r(XY) = \frac{\sum\limits_{i<j}\sum \Delta_x \Delta_y}{\sqrt{\sum\limits_{i<j}\sum \Delta_x^2 \; \sum\limits_{i<j}\sum \Delta_y^2}}$$

Het blijkt dat de korrelatie-analyse niet is gebaseerd op veranderingen binnen eenheden maar op *verschillen tussen eenheden* . Hiermee zijn strikt genomen niet langer uitspraken op individueel nivo mogelijk.

In de natuurwetenschappen kampt men met hetzelfde verschijnsel, maar hier ligt het veel minder problematisch. Stel ballon A gevuld met gas heeft bij $40^{o}C$ een volume van 42 dm^3 terwijl een met A identieke ballon B gevuld met gas van hetzelfde soort bij $60^{o}C$ een volume heeft van 45 dm^3. Wanneer nu ballon A met $20^{o}C$ wordt verhit dan mag men *onder overigens gelijkblijvende omstandigheden* aannemen dat het volume met 3 dm^3 zal stijgen. De basis waarop deze aanname berust is tweeërlei. Ten eerste kan bij natuurkundige verschijnselen de ceteris paribus klausule meestal zonder al te veel problemen worden gesteld, en voor zover dit niet het geval is kunnen de gewijzigde omstandigheden meestal voldoende gekend worden om toch tot verantwoorde kausale uitspraken te komen. Ten tweede zijn eenheden in de fysika min of meer stabiel en homogeen. Dat wil zeggen fysische objekten veranderen meestal niet sterk op korte termijn, terwijl objekten van eenzelfde soort veelal identiek zijn van vorm en samenstelling.

In de sociale realiteit ligt dit gradueel anders. Sociale eenheden veranderen voortdurend terwijl ook de omstandigheden vaak wijzigingen ondergaan. Bovendien zijn sociale eenheden dermate komplex dat strikt genomen nooit sprake kan zijn van homogeniteit. Het is duidelijk dat dit het maken van kausale gevolgtrekkingen in de sociale wetenschappen bemoeilijkt. Er zijn echter geen *principiële* redenen waarom de overgang van veranderingen binnen eenheden naar verschillen tussen eenheden niet zou kunnen worden gemaakt.

Aan het bovenstaande kunnen argumenten worden ontleend tegen het hanteren van korrelaties als startpunt van kausaal onderzoek. Er zijn echter zeker niet minder argumenten vóór te vinden. Een van de voordelen van de korrelatie-theorie is dat een stochastische komponent is toegestaan. Dit is ook nodig gezien het eerder gekonstateerde kans-aspekt van kausale relaties. Een ander voordeel is dat kontrole op schijnrelaties mogelijk is, niet alleen door middel van vergelijking van korrelatie-koëfficiënten binnen subgroepen gevormd door middel van een zogenaamde testvariabele (elaboratie), maar bovendien via het berekenen van partiële korrelatie-koëfficiënten. Een laatste punt dat in niet geringe mate pleit voor hantering van korrelaties als startpunt voor kausaal onderzoek, is gelegen in de toepassing van geavanceerde multivariate technieken. In deze studie zal een aantal van de belangrijkste daarvan aan de orde komen. (zie de hoofdstukken 2 en 3).

Tot slot een voor deze studie centrale opmerking naar aanleiding van het punt van schijnrelaties. Een schijnrelatie tussen twee variabelen ontstaat geheel óf gedeeltelijk door toedoen van externe (doorkruisende) variabelen, zodat een dergelijke relatie niet, dan wel slechts gedeeltelijk, een produktiekarakter kan worden toegekend. Terecht werd in het verleden dit opsporen van schijnbaarheid in het niet-experimentele onderzoek dan ook gezien als het belangrijkste instrument voor het onderzoeken van kausale hypothesen. Niettemin lijkt hier sprake te zijn van een eenzijdig benadrukken van schijnverbanden ten koste van aandacht voor andere problemen welke verbonden zijn aan kausaal onderzoek op basis van korrelaties. Zo moet het van groot belang worden geacht onjuistheden op te sporen welke het gevolg zijn van de waarnemings- en operationaliseringsprocedure. Kan een korrelatie berustend op schijnbaarheid in het geheel geen indikatie zijn voor het bestaan van kausaliteit, dit kan evenmin het geval zijn voor korrelaties welke zijn verstoord door meet- en operationaliseringfouten. Bovendien treedt een verstoring op in de schatting van modelparameters. Het valt niet moeilijk in te zien dat een en ander extra relevant is voor modellen met theoretische variabelen, om welke reden deze problematiek het onderwerp van studie is in de hoofdstukken 4, 5 en 6.

3. MODEL EN THEORIE

Tot nu was vrijwel uitsluitend sprake van afzonderlijke kausale relaties. Representatie van theorieën vereist evenwel de hantering van een konstellatie van dergelijke relaties. Dit betekent een overgang van enkelvoudige regressievergelijkingen naar systemen van zogenaamde strukturele vergelijkingen. Een gangbare term voor een dergelijk systeem is kausaal model of ook wel struktureel model. Deze verbreding brengt een tweetal nieuwe aspekten met zich mee, die zich laten formuleren in de volgende vragen: (a) wat moet worden verstaan onder een kausale verklaring en (b) wat is de betekenis van het modelbegrip in deze kontekst?

kausale verklaring

Een gangbare omschrijving van het begrip verklaren is het met behulp van een wet laten zien dat verschijnselen tot stand komen volgens relaties welke in die wet worden genoemd. Men spreekt hier wel van het "covering law"-type verklaring. Afhankelijk van de aard van de gehanteerde wetten zijn hierbinnen nog verschillende verklaringstypen mogelijk (zie Bunge 1961 en Pasquinelli 1967). Zo is een kausale verklaring gebaseerd op een kausale wet (-relatie) en ziet er formeel als volgt uit: X is een oorzaak van Y, welnu X doet zich voor en daarom treedt Y op. Met andere woorden, kausaal verklaren aldus opgevat staat gelijk aan afleiden uit een kausale wet.

Als nadeel van het "covering law"-type verklaring dient het gevaar vermeld te worden van te weinig aandacht voor verdergaand onderzoek naar de in wetten genoemde relaties. Strikt genomen is een statistische samenhang een relatie, maar deze te nemen als verklaringsgrond' is weinig bevredigend. Men zal juist geïnteresseerd zijn in de wijze waarop statistische samenhangen tot stand komen. Impliciet is dit ook de kern waar het bij strukturele modellen in deze studie om draait: het zoeken naar (kausale) strukturen achter korrelatie-matrices.

Stel dat een statistische samenhang wordt gekonstateerd tussen milieu en schoolprestaties bij kinderen, in die zin dat hoe hoger de skore van de ouders op de SES-schaal hoe beter de schoolprestatie van de kinderen. Aangenomen wordt dat deze samenhang kausaal kan worden geïnterpreteerd en kan worden neergelegd in een kausale wet. In zo'n geval gaat het niet aan om te zeggen dat lage schoolprestaties kunnen worden verklaard door te verwijzen naar de kausale wet hierboven. Het is de kausale wet zelf die nog nader verklaard kan worden en wel door het zoeken naar tussenschakels in het proces van veroorzaking. Een mogelijke tussenschakel in bovenstaande kausale relatie is taalvaardigheid. Hiermee kan tot een meer bevredigende kausale verklaring worden gekomen, die er als volgt uit zou kunnen zien. In lagere milieus wordt een taal gehanteerd die vaak minder goed aansluit bij de aard van de studiestof zoals bijvoorbeeld naar abstraktie-nivo en affektieve neutraliteit, waardoor veelal minder goede schoolprestaties kunnen worden geleverd. Hier blijkt het belang van het analyseren van het "tussengebeuren" waarop reeds eerder werd gewezen. Konkluderend zouden we ook kunnen zeggen dat in het "covering law"-type verklaring de nadruk teveel ligt op de variabelen in plaats van op de relaties daartussen.

Er is nog een tweede reden om bij de hantering van kausale modellen de aandacht te konsentreren op relaties in plaats van op variabelen. Vaak namelijk kan een empirische samenhang in het geheel niet kausaal worden geïnterpreteerd. Dit doet zich voor in het geval van een schijnrelatie. Stel er is ten aanzien van twee variabelen X en Y sprake van een schijnrelatie als gevolg van de variabele Z $[\rho(XY.Z) = 0]$. Ook in zo'n geval is in feite sprake van een kausale verklaring. Immers, de korrelatie tussen X en Y kan worden wegverklaard door te verwijzen naar niet één maar twéé kausale wetten. Namelijk een wet die zegt dat Z een oorzaak is van X en een wet volgens welke Z tevens een oorzaak is van Y.

Konkluderend kan worden gezegd, dat het in deze dissertatie gaat om een kausale verklaringswijze waarbij relaties tussen verschijnselen centraal staan, en niet in de eerste plaats die verschijnselen zelf. Het verschil met voorspelling is dat hier het vinden van statistische samenhangen eindprodukt is van het onderzoek terwijl in de door ons voorgestelde kausale analyse de korrelaties het startpunt vormen, waarna wordt onderzocht via welke kausale mechanismen deze korrelaties tot stand zijn gekomen.

model

Hierboven werd de term struktureel model ingevoerd. De vraag is wat hier de status is van de term model. Volgens Bertels en Nauta (1969: 28) is sprake van een model als men een bekend systeem B dat onafhankelijk is van systeem 0 benut om informatie te verkrijgen over 0. Nu stellen Diederiks en Kraan (1972: 3) dat datgene waarvan we een model wensen (het systeem 0 van Bertels en Nauta) een tweetal kategorieën van verschijnselen kan betreffen: hetzij een of ander empirisch systeem hetzij een formeel systeem ofwel theorie. Wippler (1969) maakt mijns inziens hetzelfde onderscheid waar hij spreekt over materiële respektievelijk symbolische modellen. Volgens laatstgenoemde is een symbolisch model de afbeelding van een theorie, hetgeen wil zeggen dat het hetzij isomorf hetzij analoog is met de betreffende theorie. De afbeelding ofwel formalisering komt tot stand via een symbolisch systeem zoals wiskunde, logika, programmeertaal dan wel een kombinatie hiervan. Bij de vertaling in een wiskundige taal spreekt men van mathematische modellen zoals bijvoorbeeld in het geval van stelsels strukturele vergelijkingen. De konstatering dat een struktureel model een specifieke vorm is van een mathematisch model, leidt tot de konklusie dat een struktureel model kan worden beschouwd als de representatie van een theorie. Dit is dan ook het perspektief van waaruit in deze dissertatie strukturele modellen worden bekeken.

Het onderhavige modelbegrip betreft de *vorm* van een strukturele vergelijking. Deze vorm behelst de opsomming van relevante variabelen en de aard van hun onderlinge relaties. In de kontekst van strukturele modellen slaat dit laatste op de mathematische aard van de strukturele vergelijkingen (lineaire versus kurvilineaire effekten en additieve versus multiplikatieve effekten) alsmede op de verdeling van storingstermen. Het resultaat van beslissingen op al deze punten ofwel het resultaat van "specifikatie van de vorm" is een model. In de praktijk van het sociologisch en politikologisch onderzoek betekent dit globaal dat van een specifikatiefout wordt gesproken wanneer ten onrechte tussen twee variabelen een kausale relatie dan wel het ontbreken van een relatie wordt verondersteld en wanneer ten onrechte wordt uitgegaan van samenhang dan wel het ontbreken van samenhang tussen storingstermen.

Vervolgens rijst de vraag naar het nut van het werken met modellen. Door Wippler wordt erop gewezen dat geformaliseerde theoretische konstrukties (zoals bijv. strukturele modellen) géén heuristische funktie kunnen vervullen. Het is volgens hem niet mogelijk om aan de hand van dergelijke modellen tot geheel nieuwe proposities te komen. Wel is hij van oordeel dat met de eenvoudige en semantisch inhoudsloze termen waaruit deze modellen zijn opgebouwd de achterliggende begrippen uit de theorie gemakkelijker gemanipuleerd kunnen worden, zodat een inzicht kan worden verkregen in de konsekwenties die bepaalde kombinaties van begrippen en proposities uit de theorie hebben. Om het met de woorden van Wippler zelf te zeggen (1962: 280) "Modellen in engere zin (geformaliseerde theoretische constructies) maken het mogelijk de exactheid van het denken en de duidelijkheid van formaliseringen te controleren; hun functie ligt in het verhelderen van complexe theoretische samenhangen en het verstrekken van een overzichtelijke samenvatting ervan. Sociologische theorie-vorming en definiëring van de gehanteerde begrippen moeten al voltooid zijn voordat met de constructie van dergelijke modellen kan worden begonnen". In grote lijnen deel ik deze opvatting van Wippler. Er zij echter op gewezen dat zij niet uitsluit dat aan de hand van strukturele modellen verklaring mogelijk is. Instemming verdient Land (1971), waar deze stelt dat mathematische modellen zelfs een meer algemene verklaringsbasis bieden dan verbale theorieen. Immers de modelparameters kunnen vrij variëren al naar gelang de onderzoekspopulatie, zodat een mathematisch model geheel passend kan worden gemaakt voor de populatie waarop men zijn verklaring richt. Voorts dient te worden gewezen op de toetsende funktie van mathematische modellen in het algemeen en strukturele modellen in het bijzonder. Dergelijke modellen kunnen namelijk worden opgevat als hypothesen welke via toetsing aan de empirie op hypothetisch-deduktieve wijze worden geverifiëerd. Een voordeel is dat hier meestal toetstingskriteria voorhanden zijn, dit in tegenstelling tot wat het geval is met verbale theorieën en modellen.

Een andere eenzijdigheid van de opvattingen van Wippler is dat strukturele modellen het bestaan van een theorie zouden vóóronderstellen. Zo gezien zouden strukterele modellen uitsluitend een verifikatie-funktie hebben. Hoewel dit blijkens bovenstaande zeker de hoofdzaak is, geeft dit standpunt mijns inziens aanleiding tot een onderschatting van de rol van strukturele modellen. Het is heel wel mogelijk om los van bestaande theorieën op exploratieve wijze modellen te formuleren, om vervolgens te zoeken naar verdergaande modelspecifikatie. In de literatuur zijn de laatste tijd aanzetten hiertoe te vinden.

In deze dissertatie staat centraal de ondersteuning van theorievorming middels strukturele modellen. Dit leidde tot de keuze voor modellen met theoretische variabelen, welke laatste worden gekarakteriseerd als komplexe variabelen met meestal een abstrakt en latent karakter, bij de meting waarvan behalve "random" vooral ook "nonrandom" meetfouten een rol spelen. Er zijn in het verleden vele diskussies geweest over de aard van latente variabelen en hun theoretische status. Het is in genen dele de bedoeling deze diskussie op deze plaats voort te zetten. In dit geschrift wordt uitgegaan van het standpunt dat latente variabelen, voor zover zij zijn gebaseerd op een theoretisch gefundeerde keuze van indikatoren, op zijn minst enige theoretische relevantie moet worden toegekend.

2. STRUKTURELE MODELLEN

Dat korrelaties onder bepaalde kondities geschikt zijn als basis voor kausale analyse, betekent bij het gebruik van strukturele modellen dat gezocht wordt naar kausale strukturen achter konfiguraties van korrelaties, welke strukturen worden vastgelegd in strukturele parameters. In mathematisch-statistische zin houdt het voorgaande in de schatting van een relatief klein aantal parameters, waarmee zo goed mogelijk de korrelatiematrix betreffende samenhangen tussen de modelvariabelen kan worden gereproduceerd. Het objekt van de volgende hoofdstukken nù is het verkrijgen van betrouwbare en geldige schattingen van parameters in modellen met ongemeten ofwel theoretische variabelen.

In het onderhavige hoofdstuk volgt een eerste aanzet daartoe. In paragraaf 1 wordt nagegaan in hoeverre en op welke wijze regressie-analyse geschikt is voor het onderzoeken van kausale hypothesen. Vervolgens wordt onderzocht of en in hoeverre systemen van vergelijkingen geschikt zijn voor analyse van konstellaties van kausale hypothesen, dat wil zeggen kausale modellen. Dergelijke modellen worden reeds geruime tijd onderzocht en gebruikt door ekonometristen onder de benaming "simultaneous equation systems" en "structural equation models". Binnen de kategorie van strukturele modellen zijn een tweetal subkategorieën te onderkennen te weten niet-rekursieve modellen en, als een bijzonder geval hiervan, rekursieve modellen. Beide komen uitvoerig aan de orde in de respektievelijke paragrafen 2 en 3.

1. REGRESSIE-ANALYSE VAN KAUSALE RELATIES

Bij een funktie y=f(x) hangen x en y volledig met elkaar samen, zodanig dat associatiekoëfficiënten hier hun maximale waarde gelijk aan één bereiken. Het lijkt dan ook dat het begrip funktie via het idee van samenhang aansluiting geeft op het in hoofdstuk 1 uitgewerkte kausaliteitsbegrip. In datzelfde hoofdstuk kwamen evenwel diverse argumenten naar voren voor hantering van een kausaliteitsbegrip met een kanskarakter. Dit leidt tot een overstap naar de regressietheorie. Zoals bekend is, kan een regressievergelijking worden opgevat als een funktievergelijking waaraan een stochastische komponent is toegevoegd. Deze laatste wordt aangeduid als de storingsterm. Behalve het kanskarakter kunnen ook de overige in hoofdstuk 1 genoemde elementen van een kausale relatie in een regressievergelijking tot hun recht komen. Zo wordt de sterkte van kausale afhankelijkheid uitgedrukt in een regressiekoëfficiënt, wordt de asymmetrie verdiskonteerd in de keuze van een afhankelijke en een onafhankelijke variabele en komt de valentie tot uiting in het teken van de regressiekoëfficiënt. Met de tijdsvolgorde kan uiteraard rekening worden gehouden door de afhankelijke variabele niet op een later tijdstip te meten dan de onafhankelijke variabele. Ook indirekte kausale relaties kunnen worden onderzocht, hoewel daarvoor een uitbreiding van de regressietheorie is vereist (zie de paragrafen 2 en 3).

Met het oog op een nadere beschouwing van de mogelijkheid van regressie-analyse als hulpmiddel bij kausaal onderzoek, wordt onderscheid gemaakt tussen regressie als schatting of voorspelling en regressie als kausaal-analytisch instrument. In het laatste geval spreek ik voorlopig over kausale regressie-analyse, een term die verderop weer zal worden vervangen. Formeel wijken beide toepassingen niet van elkaar af, maar er is een verschil in terminologie, praktische onderzoeksaanpak en interpretatie van de analyseresultaten. Zo kan men in de kausale regressie-analyse beter niet spreken van regressoren en regressanten maar van respektievelijk onafhankelijke en afhankelijke variabelen, om aan te geven dat genoemde typen variabelen *worden geïnterpreteerd* als respektievelijk oorzaken en gevolgen. Wat betreft de praktische onderzoeksaanpak denk ik aan de aard en het aantal van de te selekteren onafhankelijke variabelen en aan de vraag of in de analyse de nadruk moet liggen op modelparameters dan wel op regressie-residuen. Een en ander wordt hieronder nader uitgewerkt.

Parallel aan het onderscheid tussen regressie als schatting enerzijds en kausale analyse anderzijds kan men in de regressie-analyse een tweetal doelen nastreven: (a) het minimaliseren van regressie-residuen en (b) het zoeken naar betrouwbare en geldige regressiekoëfficiënten. In het eerste geval spreekt men van schatting of voorspelling. De aandacht is gericht op de regressie-residuen ofwel schattingsfouten. Hoe geringer de schattingsfout hoe hoger de koëfficiënt van multiple determinatie als indikator van voorspelbaarheid van de regressant op basis van regressoren. Invoering van als maar meer regressoren zal deze koëfficiënt doen toenemen, tenzij toegevoegde regressoren perfekt samenhangen met een of meerdere reeds in de vergelijking aanwezige regressoren. Er worden geen eisen gesteld aan de aard en het aantal van de te selekteren regressoren, waaruit het ad hoc karakter van de voorspelling blijkt. Het enig geldende kriterium is minimalisering van de schattingsfout. In het tweede geval, te weten dat van de kausale regressie-analyse, gaat de interesse veel meer uit naar een juiste specifikatie van de regressievergelijking en, in het verlengde hiervan, naar het verkrijgen van geldige en betrouwbare regressiekoëfficiënten.

Er zijn twee probleemgebieden die dit laatste in de weg staan en die vallen buiten de traditionele kriteria van zuiverheid, efficiëntie en konsistentie van schatting. Deze betreffen de problematiek van *meetfouten* en, als veralgemening hiervan, die van ongemeten variabelen alsmede die van *specifikatiefouten* .In de ekonometrie staan deze bekend als respektievelijk "errors in variables" en "errors in equations". Om meer inzicht te krijgen in de vraag hoe deze een betrouwbare en geldige schatting van parameters in de weg staan, volgt hieronder een uitwerking van beide probleemgebieden.

specifikatiefouten

Onder de term specifikatiefouten vallen alle onjuistheden die optreden wanneer de vorm en de samenstelling van een regressievergelijking niet in overeenstemming zijn met de achterliggende processen in de realiteit. De vorm van een regressievergelijking betreft onder andere de vraag of het gaat om additieve dan wel multiplikatieve en lineaire dan wel kurvilineaire effekten. De samenstelling heeft betrekking op de vraag of bepaalde variabelen terecht in een vergelijking zijn opgenomen dan wel eruit zijn weggelaten. Wanneer op genoemde punten foutieve beslissingen worden genomen, heeft dit konsekwenties voor de parameterschatting. Het meest voorkomende geval van specifikatiefout is het ten onrechte weglaten van een of meer variabelen uit een vergelijking. Uitgangspunt is de vergelijking (variabelen in afwijkingskores dat wil zeggen in afwijking van de populatie-gemiddelden).

(2.1) $$y = \beta_1 x_1 + \beta_2 x_2 + u$$

Deze vergelijking beschrijft het mechanisme volgens welk x_1 en x_2 tesamen y bepalen. Storingsterm u wordt geacht onafhankelijk te zijn van x_1 en x_2. Stel nu dat de onderzoeker onkundig is van het feit dat x_2 naast x_1 de waarde van y bepaalt en ten onrechte uitgaat van de vergelijking

(2.2) $$y = \beta_1^{\square} x_1 + u^{\square}$$

waar de hokjes aangeven dat het gaat om de gewijzigde situatie zonder x_2. De vraag is of er een verschil is tussen β_1 en $\beta_1^{\square}$ alsook tussen u en $u^{\square}$. Voor de populatieregressie van y op x_1 geldt

(2.3) $$E(y|x_1) = (\sigma(x_1 y)/\sigma^2(x_1))x_1$$

waar E staat voor mathematische verwachtingswaarde. Voor $\sigma(x_1 y)$geldt naar aanleiding van (2.1)

$$\sigma(x_1 y) = \beta_1 \sigma^2(x_1) + \beta_2 \sigma(x_1 x_2)$$

Substitutie in (2.3) geeft

$$E(y|x_1) = \beta_1^{\square} x_1 = (\beta_1 + k\beta_2)x_1$$

waar $k = \sigma(x_1 x_2)/\sigma^2(x_1)$, zijnde de populatie-regressiekoëfficiënt bij regressie van x_2 op x_1. Het blijkt dat weglating van x_2 uit de vergelijking leidt tot een "opwaardering" van β_1 ter grootte $k\beta_2$. In engelstalige literatuur spreekt men hier van "amplification". Deze verstoring blijft evenwel uit wanneer x_1 en x_2 niet met elkaar samenhangen aangezien in dat geval geldt $k=0$. In het vervolg van deze studie wordt hierop niet verder ingegaan om zodoende meer plaats te maken voor een behandeling van de meetfoutenproblematiek.

meetfouten

Hoewel een uitgebreide behandeling van de meetfoutenproblematiek wordt uitgesteld tot de hoofdstukken 3, 4, 5 en 6, volgt hieronder een korte schets van de verstoring welke *random* meetfouten teweegbrengen bij het schatten van modelparameters. Bekend is dat deze meetfouten in de afhankelijke variabele in een regressievergelijking de regressiekoëfficiënten niet verstoren, terwijl dit wel het geval is wanneer de onafhankelijke variabelen met meetfouten zijn gemeten. Uitgangspunt is de volgende regressievergelijking (variabelen in afwijkingsskores)

$$y = \beta x^* + u$$

waar het sterretje aangeeft dat het gaat om een variabele vrij van random meetfouten en waar x^* en u niet samenhangen. Stel voorts dat x aangeeft de variabele welke resulteert uit de meting van x^*. Deze meting kan via een zogenaamde meetvergelijking worden weergegeven

$$x = x^* + e$$

waar e aangeeft een ongemeten variabele van random meetfouten welke geacht wordt niet samen te hangen met x^* en u.

Voor x en y gelden de volgende populatievarianties en -kovarianties

$$\sigma^2(x) = \sigma^2(x^*) + \sigma^2(e), \sigma(xy) = \beta\sigma^2(x^*), \sigma^2(y) = \beta^2\sigma^2(x^*) + \sigma^2(u)$$

zodat voor de populatieregressie van y op x kan worden geschreven

$$E(y|x) = (\sigma(xy)/\sigma^2(x))x = \beta kx$$

waar

$$k = \sigma^2(x^*)/(\sigma^2(x^*) + \sigma^2(e))$$

De aldus gedefinieerde grootheid k kan, zoals in hoofdstuk 5 blijkt, worden beschouwd als een betrouwbaarheidskoëfficiënt. De waarde ervan ligt tussen 0 en 1, waaruit blijkt dat meetfouten leiden tot een onderschatting van de nulde-orde regressiekoëfficiënt. In engelstalige literatuur spreekt men hier van "attenuation". Met behulp van matrix-algebra kan een en ander op waarnemingsnivo worden uitgewerkt voor een vergelijking met H onafhankelijke variabelen. De skores op deze variabelen worden verzameld in een $N*H$ matrix $\underset{\sim}{X}$. Daarnaast wordt een $N*H$ matrix $\underset{\sim}{X}^*$ gedefinieerd met errorvrije skores op deze variabelen. De verstoring van het meetproces kan worden weergegeven in de volgende meetvergelijking

$$(2.4) \qquad \underset{\sim}{X} = \underset{\sim}{X}^* + \underset{\sim}{E}$$

waar $\underset{\sim}{E}$ een $N*H$ matrix is met meetfouten e. Voorts veronderstellen we

$$(2.5) \qquad \underset{\sim}{y} = \underset{\sim}{X}^*\underset{\sim}{\beta} + \underset{\sim}{u}$$

waar $\underset{\sim}{y}$ is een $N*1$ vektor met skores op de afhankelijke variabele, $\underset{\sim}{\beta}$ is een $H*1$ vektor met parameters en $\underset{\sim}{u}$ is een

N $*1$ vektor met storingstermen. Substitutie van (2.4) in (2.5) geeft

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + (\underset{\sim}{u} - \underset{\sim}{E}\,\underset{\sim}{\beta})$$

De OLS schatter voor deze vergelijking ziet er als volgt uit

$$\hat{\underset{\sim}{\beta}} = \underset{\sim}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'(\underset{\sim}{u} - \underset{\sim}{E}\underset{\sim}{\beta})$$

Het valt niet te verwachten dat met behulp van OLS een zuivere en konsistente schatting van $\underset{\sim}{\beta}$ kan worden verkregen. Dit blijkt wanneer $[\underset{\sim}{X}'(\underset{\sim}{u}-\underset{\sim}{E}\,\underset{\sim}{\beta})]$ nader wordt onderzocht.

$$\text{plim}(\underset{\sim}{X}'(\underset{\sim}{u} - \underset{\sim}{E}\beta)) = \text{plim}(\underset{\sim}{X}'\underset{\sim}{u}) - \text{plim}(\underset{\sim}{X}'\underset{\sim}{E})\underset{\sim}{\beta}$$

Veelal mag worden aangenomen dat de storingsterm $\underset{\sim}{u}$ niet samenhangt met $\underset{\sim}{X}$ zodat

$$\text{plim}(\underset{\sim}{X}'\underset{\sim}{u}) = \underset{\sim}{0} \tag{2.6}$$

Voorts geldt

$$\text{plim}(\underset{\sim}{X}'\underset{\sim}{E}) = \text{plim}(\underset{\sim}{X}'\underset{\sim}{E}) + \text{plim}(\underset{\sim}{E}'\underset{\sim}{E}) \tag{2.7}$$

Het valt niet te verwachten dat deze vergelijking de waarde nul aanneemt, zelfs niet wanneer de meetfouten alle statistisch onafhankelijk zijn van de errorvrije variabelen (zie de eerste term in het rechterdeel van (2.7)). Immers in de tweede term van het rechterdeel staat de variantie-kovariantie-matrix van meetfouten, welke onder normale kondities ongelijk is aan de nulmatrix. Het blijkt dus dat $\text{plim}(\hat{\underset{\sim}{\beta}})$ hierboven ongelijk is aan $\underset{\sim}{\beta}$.

We kunnen dit resultaat ook aldus zien. Indien we geen rekening zouden houden met meetfouten en een OLS-schatting zouden uitvoeren op de vergelijking

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{u}$$

dan zouden de meetfouten terecht komen in de storingstermen. Dit zou betekenen dat als (2.7) ongelijk is aan nul dat dan ook (2.6) niet langer opgaat. Met andere woorden, ook in het geval van meetfouten treedt er een korrelatie op tussen storingsterm u en de onafhankelijke variabelen uit de vergelijking.

Een laatste voor de handliggend gevolg van meetfouten, dat van centrale betekenis zal blijken te zijn voor deze studie, is een samenhang tussen de storingstermen uit twee verschillende modelvergelijkingen. Deze samenhang doet zich voor wanneer twee vergelijkingen beide dezelfde met meetfout gemeten regressor(en) bevatten. Het is duidelijk dat genoemde samenhang het gevolg is van de meetfout(en) die beide storingstermen als gemeenschappelijk element hebben.

Deze paragraaf resumerend kan worden gekonkludeerd dat bij toepassing van de regressietheorie als instrument van kausale analyse, de regressiekoëfficiënt en niet de (gekwadrateerde) korrelatiekoëfficiënt, dan wel derivaten daarvan zoals de F-toets, centraal staat in tegenstelling tot wat het geval is bij regressie als instrument van voorspelling. Dit betekent onder andere dat specifikatie- en meetfouten een bijzondere aandacht behoeven, wil men kunnen komen tot betrouwbare en geldige kennis van kausale relaties. Het bleek dat twee van de belangrijkste assumpties van de regressie-analyse, te weten het ontbreken van samenhang tussen regressoren en storingstermen, alsook tussen storingstermen onderling, onhoudbaar zijn wanneer genoemde foutenbronnen zich voordoen. Hiermee loopt de schatting van regressiekoëfficiënten uit de hand met als gevolg foutieve konklusies ten aanzien van de schijnbaarheid en de sterkte van kausale afhankelijkheden.

2. SYSTEMEN VAN VERGELIJKINGEN

Kenmerkend voor kausaal onderzoek is meestal dat men te maken heeft met meerdere afhankelijke variabelen, voor elke waarvan een modelvergelijking wordt uitgeschreven. Het feit dat meerdere vergelijkingen in de analyse zijn betrokken, geeft zoals zal blijken zodanige komplikaties dat moet worden gesproken van een systeem van interdependente vergelijkingen. In de ekonometrie noemt men modellen bestaande uit een verzameling vergelijkingen waarin kausale relaties worden weergegeven, strukturele modellen en de afzonderlijke vergelijkingen strukturele vergelijkingen. De term regressiekoëfficiënt wordt vervangen door strukturele parameter.

Er is een goede reden waarom in de titel hierboven niet wordt gesproken van regressievergelijkingen. In een regressievergelijking kunnen namelijk uitsluitend *direkte* kausale relaties worden gerepresenteerd. In strukturele modellen komen evenwel veelvuldig indirekte relaties voor. Het gevolg is dat er meestal veel minder dan L(L-1) modelparameters moeten worden geschat, waar L staat voor het aantal variabelen in een struktureel model. Hierdoor ontstaat een aspekt dat niet aanwezig is in regressievergelijkingen, namelijk dat een groot aantal korrelaties wordt verklaard door een relatief klein aantal kausale relaties, weergegeven door modelparameters.

notatie en assumpties

Beschouwen we een struktureel model met G lineaire vergelijkingen in L variabelen en G ongemeten storingstermen voor een steekproef van N eenheden.

$$\underset{\sim}{A}\underset{\sim}{z}_n = \underset{\sim}{u}_n \qquad n=1,\ldots,N \tag{2.8}$$

waar

$$\underset{\sim}{A} = \begin{pmatrix} a_{11} & \cdots & a_{1L} \\ \vdots & & \vdots \\ a_{G1} & \cdots & a_{GL} \end{pmatrix}, \quad \underset{\sim}{z}_n = \begin{pmatrix} z_{n1} \\ \vdots \\ z_{nL} \end{pmatrix}, \quad \underset{\sim}{u}_n = \begin{pmatrix} u_{n1} \\ \vdots \\ u_{nG} \end{pmatrix}$$

$\underset{\sim}{A}$ is een G*L matrix met strukturele parameters of ook wel modelparameters, $\underset{\sim}{z}_n$ is een L*1 vektor met gemeten variabelen en $\underset{\sim}{u}_n$ is een G * 1 vektor met storingstermen. Voorts dient een onderscheid te worden gemaakt tussen endogene en exogene variabelen. Als exogeen worden variabelen beschouwd welke uitsluitend kausaal afhankelijk zijn van variabelen buiten het model. Essentiëel voor de omschrijving van een exogene variabele is verder dat hij statistisch onafhankelijk is van alle storingstermen uit het model. Endogeen zijn variabelen welke door variabelen uit het model, hetzij exogeen hetzij zélf endogeen, kausaal worden beïnvloed.

Stel er is voor elke eenheid n een verzameling endogene variabelen y_g g=1,....,G, een verzameling exogene variabelen x_h h=1,....,H (H=L-G) en een verzameling storingstermen u_g g=1,....,G, verzameld in de respektievelijke vektoren $\underset{\sim}{y}_n$, $\underset{\sim}{x}_n$ en $\underset{\sim}{u}_n$. De strukturele parameters behorende bij de endogene variabelen duiden we aan met β en die bij exogene variabelen met γ, welke parameters worden verzameld in de respektievelijke matrices $\underset{\sim}{B}$ en $\underset{\sim}{\Gamma}$. Op basis van bovenstaande worden matrix $\underset{\sim}{A}$ en vektor $\underset{\sim}{z}_n$ in (2.8) als volgt gepartitioneerd.

$$\underset{\sim}{A} = [\underset{\sim}{B} \quad \underset{\sim}{\Gamma}] \qquad \text{en} \quad \underset{\sim}{z}_n = \begin{bmatrix} \underset{\sim}{y}_n \\ \\ \underset{\sim}{x}_n \end{bmatrix} \quad n=1,....,N$$

waar

$$\underset{\sim}{B} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1G} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \beta_{G1} & \cdots & \beta_{GG} \end{pmatrix}, \; \underset{\sim}{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1H} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \gamma_{G1} & \cdots & \gamma_{GH} \end{pmatrix}, \; \underset{\sim}{y}_n = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_G \end{pmatrix}, \; \underset{\sim}{x}_n = \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_H \end{pmatrix}$$

Vergelijking (2.8) gaat nu over in

(2.9) $$\underset{\sim}{B}\,\underset{\sim}{y}_n + \underset{\sim}{\Gamma}\,\underset{\sim}{x}_n = \underset{\sim}{u}_n \quad n=1,\ldots,N$$

Deze vergelijkingen tesamen noemt men wel de strukturele vorm van het model. Om deze reden worden de storingstermen u wel strukturele storingstermen genoemd. Wanneer elke endogene variabele wordt geschreven in termen van uitsluitend exogene variabelen, verkrijgen we de zogenaamde gereduceerde vorm van het model. Anders gezegd, de gereduceerde vorm is de oplossing van de strukturele vorm voor $\underset{\sim}{y}_n$.

(2.10) $$\underset{\sim}{y}_n = \underset{\sim}{\Pi}\,\underset{\sim}{x}_n + \underset{\sim}{v}_n \quad n=1,\ldots,N$$

waar

$$\underset{\sim}{\Pi} = \underset{\sim}{B}^{-1}\,\underset{\sim}{\Gamma} \quad \text{en} \quad \underset{\sim}{v}_n = \underset{\sim}{B}^{-1}\,\underset{\sim}{u}_n$$

$\underset{\sim}{\Pi}$ is de G*H matrix met gereduceerde-vorm-parameters en $\underset{\sim}{v}_n$ is de G*1 vektor met gereduceerde-vorm-storingstermen.

Vervolgens wordt nagegaan welke de assumpties zijn die bij de specifikatie van strukturele modellen moeten worden gesteld met het oog op de identifikatie en schatting van strukturele parameters. Allereerst de assumpties ten aanzien van de mathematische vorm, te weten additiviteit en lineairiteit. De laatste kunnen we kontroleren aan de hand van een puntenwolk-analyse, geheel ten onrechte in de praktijk zelden gedaan. Een nadeel is wel dat de lineairiteit aldus bivariaat wordt bekeken, terwijl strukturele modellen multivariaat zijn van karakter. Voor een kontrole op de additiviteit zijn bij mijn weten geen statistische methodieken uitgewerkt, zodat op dit punt moet worden vertrouwd op a priori inzichten en ad hoc procedures. Alle overige te stellen assumpties kunnen worden onderscheiden in twee stellen, te weten a) een stel met stochastische assumpties betrekking hebbend op de storingstermen, ook wel genoemd de impliciete variabelen, en b) een stel met assumpties betreffende alle endogene en exogene variabelen, ook wel genoemd de expliciete variabelen.

stochastische assumpties

Ten aanzien van de strukturele vorm in (2.9) worden meestal de volgende assumpties gesteld:

a_1. $E(\underset{\sim}{u}_n) = \underset{\sim}{0}_n$ $\quad$ n=1,.....,N

a_2. $E(\underset{\sim}{x}_n \underset{\sim}{u}'_n) = \underset{\sim}{0}_n$ $\quad$ n=1,.....,N

a_3. $E(\underset{\sim}{u}_n \underset{\sim}{u}'_m) = \underset{\sim}{\Sigma}$ $\quad$ n,m=1,.....,N; n = m

a_4. $E(\underset{\sim}{u}_n \underset{\sim}{u}'_m) = \underset{\sim}{0}$ $\quad$ n,m=1,....,N; n $\neq$ m

In a_1 staat dat de storingstermen in alle vergelijkingen voor alle eenheden een verwachtingswaarde gelijk aan nul hebben en in a_2 dat de storingstermen niet samenhangen met de exogene variabelen. Punt a_3 behelst de aanname van homoscedasticiteit, althans voor wat betreft de elementen in de hoofddiagonaal van $\underset{\sim}{\Sigma}$. De overige elementen in $\underset{\sim}{\Sigma}$ betreffen kovarianties tussen storingstermen uit verschillende vergelijkingen, de zogenaamde "contemporaneous covariances".[1)] Punt a_4 tenslotte behelst de aanname van afwezigheid van autokorrelatie. Zoals zal blijken, speelt de variantie-kovariantie matrix van storingstermen ($\underset{\sim}{\Sigma}$) een belangrijke rol in de problemen van identifikatie en schatting van strukturele modellen. Deze matrix ziet er als volgt uit

$$\underset{\sim}{\Sigma} = E(\underset{\sim}{u}_n \underset{\sim}{u}'_n) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1G} \\ \vdots & & \vdots \\ \sigma_{G1} & \cdots & \sigma_{GG} \end{pmatrix} \quad n=1,..,N \tag{2.11}$$

Zoals elke kovariantie-matrix is $\underset{\sim}{\Sigma}$ symmetrisch positief semi-definiet (zie Christ 1966: 641-643) en is in het algemeen niet diagonaal.

De vraag rijst of de gestelde assumpties eveneens gelden voor de gereduceerde vorm in (2.10).

ad a_1. $E(\underset{\sim}{v}_n) = \underset{\sim}{B}^{-1} E(\underset{\sim}{u}_n) = \underset{\sim}{0}_n \quad n=1,..,N$

ad a_2. $E(\underset{\sim}{x}_n \underset{\sim}{v}'_n) = \underset{\sim}{B}^{-1} E(\underset{\sim}{x}_n \underset{\sim}{u}'_n) = \underset{\sim}{0}_n \quad n=1,..,N$

ad a_3. $E(\underset{\sim}{v}_n \underset{\sim}{v}'_m) = \underset{\sim}{B}^{-1} E(\underset{\sim}{u}_n \underset{\sim}{u}'_m) \underset{\sim}{B}'^{-1} = \underset{\sim}{B}^{-1} \underset{\sim}{\Sigma} \underset{\sim}{B}'^{-1} = \underset{\sim}{\Omega} \quad n,m=1,..,N;\ n=m$

ad a_4. $E(\underset{\sim}{v}_n \underset{\sim}{v}'_m) = \underset{\sim}{B}^{-1} E(\underset{\sim}{u}_n \underset{\sim}{u}'_m) \underset{\sim}{B}'^{-1} = \underset{\sim}{0}_n \quad n,m=1,..,N;\ n\neq m$

De konklusie is dat de stochastische assumpties a_1 t/m a_4 eveneens opgaan voor de gereduceerde vorm. Dit resultaat was te verwachten aangezien $\underset{\sim}{v}_n$ een lineaire funktie is van $\underset{\sim}{u}_n$.

assumpties betreffende expliciete variabelen

Definiëren we $\underset{\sim}{X}$ als de $G*N$ observatiematrix met skores op de exogene variabelen. De genoemde assumpties zijn dan

b_1. De elementen van $\underset{\sim}{X}$ hebben een gefixeerde waarde.
b_2. $\underset{\sim}{X}$ heeft een rang gelijk aan H.
b_3. Alle expliciete variabelen zijn minstens van intervalnivo.
b_4. Alle exogene variabelen zijn betrouwbaar gemeten.
b_5. De matrix $\underset{\sim}{B}$ heeft rang G.

Assumptie b_1 houdt in dat bij herhaalde steekproeftrekking de variantie in storingsterm $\underset{\sim}{u}$ de enige bron is van variantie in de afhankelijke variabele $\underset{\sim}{y}$. Assumptie b_2 zegt dat er geen exakte lineaire relaties bestaan tussen de exogene variabelen (perfekte multikolineairiteit). Met b_3 wordt aangenomen dat de expliciete variabelen met een vaste meeteenheid zijn gemeten. De vierde assumptie is nodig om te voorkomen dat meetfouten een storende invloed hebben op de parameterschatting en de laatste assumptie is nodig wil het model oplosbaar zijn voor $\underset{\sim}{y}_n$ in termen van $\underset{\sim}{x}_n$ en $\underset{\sim}{u}_n$. Immers hiervoor is nodig dat de inverse $\underset{\sim}{B}^{-1}$ kan worden berekend.

endogene regressoren

Een van de belangrijkste assumpties uit de regressietheorie is de afwezigheid van samenhang tussen regressor en storingsterm. Voor exogene variabelen gaat deze assumptie per definitie op. De vraag is of dit ook geldt voor endogene variabelen. Deze vraag is van belang omdat in strukturele modellen vaak variabelen voorkomen die in de ene vergelijking optreden als afhankelijke en in de andere als onafhankelijke variabele. Dergelijke variabelen worden voortaan aangeduid als "endogene regressoren", althans in hun hoedanigheid van onafhankelijke variabele (regressor). Teneinde in de gestelde vraag enig inzicht te krijgen, ga ik uit van het volgende niet-rekursieve model

$$y_1 = \beta_1 y_2 + u_1$$

$$y_2 = \beta_2 y_1 + u_2$$

Dit model is weliswaar ondergeïdentificeerd maar leent zich niettemin voor adstruktie van de hierboven gestelde vraag. Bekijken we de eventuele samenhang tussen y_1 en u_2. Substitutie van de tweede vergelijking in de eerste geeft

$$y_1 = \beta_1 \beta_2 y_1 + \beta_1 u_2 + u_1$$

ofwel

$$y_1 = [\beta_1 / (1 - \beta_1 \beta_2)] \quad u_2 \quad + [1/(1 - \beta_1 \beta_2)] \, u_1$$

Vermenigvuldiging met u_2 geeft na het nemen van verwachtingswaarden

$$(2.12) \quad E(y_1 u_2) = [\beta_1 / (1-\beta_1 \beta_2)] E(u_2^2) + [1/(1-\beta_1 \beta_2)] \, E(u_1 u_2)$$

Het blijkt dat in het geval van symmetrische kausale relaties (niet-rekursiviteit) een samenhang optreedt tussen endogene regressor en storingsterm. Zoals (2.12) laat zien, geldt dit ongeacht de vraag of er sprake is van "contemporaneous" kovarianties tussen storingstermen. In een volgende paragraaf zal blijken dat voor *rekursieve* modellen uitsluitend een samenhang tussen endogene regressor en storingsterm kan optreden wanneer de storingstermen "contemporaneous" samenhangen.

Een en ander kan worden uitgewerkt voor een algemeen struktureel model zoals weergegeven in (2.9) en (2.10). Postmultiplikatie van de gereduceerde vorm in (2.10) met $\underset{\sim}{u}_n'$ geeft na het nemen van verwachtingswaarden

$$(2.13) \quad \underset{\sim}{\Sigma}_{yu} = E(\underset{\sim}{y}_n \underset{\sim}{u}_n') = E\,[(\underset{\sim}{\Pi} \underset{\sim}{x}_n + \underset{\sim}{v}_n) \underset{\sim}{u}_n'] = \underset{\sim}{\Pi} E(\underset{\sim}{x}_n \underset{\sim}{u}_n') + E(\underset{\sim}{v}_n \underset{\sim}{u}_n')$$

$$= \underset{\sim}{0}_n + E(\underset{\sim}{B}^{-1} \underset{\sim}{u}_n \underset{\sim}{u}_n') = \underset{\sim}{B}^{-1} \underset{\sim}{\Sigma}$$

In het algemene strukturele model worden geen restrikties opgelegd aan $\underset{\sim}{B}$ en $\underset{\sim}{\Sigma}$ zodat de elementen in $\underset{\sim}{\Sigma}_{yu}$ ongelijk aan nul zullen zijn. Met andere woorden, normaliter is er een samenhang tussen endogene regressoren en storingsterm. Dit betekent dat we ten aanzien van (2.9) de endogene variabelen niet telkens alle op één na naar de rechterkant van de vergelijking kunnen brengen om de aldus geformuleerde vergelijking te behandelen met de normale kleinste kwadraten procedure (OLS), aangezien dit zou leiden tot onzuivere en inkonsistente schatters. (Zie bijv. Johnston 1972, 351). Men is hier aangewezen op andere schattingsmethoden zoals "maximum likelyhood" methoden of, indien er voldoende overidentifikatie is in het model, een "two stage least squares"- dan wel "instrumentele variabelen"- procedure.[2)]

Samenvattend kan het volgende worden gezegd. In de vorige paragraaf bleek een samenhang te kunnen optreden tussen regressor en storingsterm als gevolg van meet- en of specifikatiefouten. Thans blijkt dat hetzelfde kan gebeuren in een systeem van regressievergelijkingen waarin endogene regressoren voorkomen. Dit betekent dat een schatting per afzonderlijke vergelijking zou leiden tot inkonsistente schattingen. Bovendien ontstaat, in het geval een en dezelfde endogene regressor in meerdere vergelijkingen aanwezig is, kans op "contemporaneous" kovarianties tussen storingstermen. Dit betekent dat de vergelijkingen in een struktureel model niet los staan van elkaar maar moeten worden gezien als een interdependent systeem.

identifikatie

Bij de analyse van strukturele modellen dient onderscheid te worden gemaakt naar een tweetal probleemgebieden, te weten de identifikatie en schatting van strukturele parameters. Beide geven een zeer breed terrein van vooral ekonometrisch onderzoek aan. Het is niet de bedoeling een overzicht te geven van dit gebied. Volstaan wordt met een korte schets van het identifikatieprobleem. Wat betreft de schattingsproblematiek worden enkele schattingsprocedures genoemd, welke met het oog op in verdere hoofdstukken te behandelen modellen met ongemeten variabelen speciaal van belang zijn.

Een bekend methodologisch probleem doet zich voor wanneer meerdere theorieën alle een even goede verklaring geven van een bepaald verschijnsel. Het problematische is in zo'n geval dat empirische data ons niet in staat stellen tot het doen van een keuze uit een van deze theorieën. Ditzelfde probleem nu doet zich voor bij strukturele modellen. Ter verduidelijking wordt onderscheid gemaakt tussen model en struktuur. Een *model* is de resultante van de specifikatie van vorm en samenstelling van vergelijkingen en van de verdeling van storingstermen. Deze specifikatie blijkt goeddeels uit de samenstelling van $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$ en $\underset{\sim}{\Sigma}$. Het betreft hier a priori dan wel theoretische kennis, bijvoorbeeld die aangaande de vraag of sommige variabelen wel of niet zijn opgenomen in een bepaalde vergelijking. Ook is het mogelijk dat vooraf kennis bestaat ten aanzien van de onderlinge verhoudingen tussen elementen in $\underset{\sim}{B}$ en $\underset{\sim}{\Gamma}$. Ten aanzien van $\underset{\sim}{\Sigma}$ is de assumptie denkbaar dat dit een diagonale matrix is, hetgeen inhoudt afwezigheid van "contemporaneous" kovariantie. Een struktuur ontstaat uit een model wanneer numerieke waarden worden toegekend aan $\underset{\sim}{B}$ $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Sigma}$. Ervan uitgaande dat de specifikatie van het model juist is, geldt in het algemeen de veronderstelling dat één bepaalde struktuur behorende bij dit model de empirische data genereert welke in het strukturele model worden gerepresenteerd.

Analoog aan het hierboven geschetste probleem bestaat nu de mogelijkheid dat twee verschillende strukturen uit één en hetzelfde model exakt dezelfde konsekwenties hebben voor de empirische data. Zulke strukturen noemt men wel "observationally equivalent" en het probleem dat eruit voortvloeit het identifikatieprobleem. Dit probleem gaat logisch vooraf aan het schattingsprobleem. Immers, wanneer op grond van de data de modelparameters worden geschat, dan weet men niet of de parameters uit de ene, dan wel die uit de andere ("observationally equivalent") struktuur worden geschat. Dit betekent dat slechts die parameters zijn geïdentificeerd welke in alle "observationally equivalent" strukturen een identieke waarde hebben.

Het valt niet moeilijk in te zien dat die strukturen "observationally equivalent" zijn, welke een identieke gereduceerde vorm hebben. Met andere woorden, het identifikatieprobleem doet zich voor wanneer meer dan één strukturele vorm uit een gereduceerde vorm afgeleid kan worden. De kans dat een strukturele vorm op unieke wijze uit een gereduceerde vorm kan worden afgeleid, is niet erg groot. Immers er zijn in $\underset{\sim}{\Pi}$ G*H elementen aanwezig terwijl in $\underset{\sim}{B}$ en $\underset{\sim}{\Gamma}$ tesamen [(G*G) + (H*G)] elementen aanwezig zijn. Meestal is het dan ook nodig dat restrikties worden opgelegd aan $\underset{\sim}{B}$ en $\underset{\sim}{\Gamma}$. Deze restrikties kunnen in principe een tweetal vormen aannemen:

a) bepaalde parameters in $\underset{\sim}{B}$ en $\underset{\sim}{\Gamma}$ hebben een a priori waarde. Wanneer deze a priori waarde gelijk is aan nul spreekt men van eksklusie-restrikties.
b) bepaalde parameters vertonen een bepaalde ratio, waartoe ook kan behoren de ratio gelijk aan één.

In de sociologie beperkt men zich steeds tot eksklusie-restrikties (zie bijvoorbeeld Boudon 1970 en Land 1973 alsook de volgende paragraaf van dit hoofdstuk). Zo zijn de in deze discipline gangbare rekursieve modellen gebaseerd op de aanname " $\underset{\sim}{\Sigma}$ diagonaal" en "$\underset{\sim}{B}$ driehoekig" (zie paragraaf 3). Het kan worden bewezen dat onder deze kondities steeds de strukturele vorm op unieke wijze uit de gereduceerde vorm kan worden afgeleid. Verdere informatie betreffende het identifikatie-probleem vindt men o.a. in Christ 1966, Johnston 1972 en Land 1973.

schatting

Aangaande de schatting van modelparameters kan onderscheid worden gemaakt tussen twee kategorieën van schattingstechnieken: (a) "single-equation" technieken en (b) "full- information" technieken. In het eerstgenoemde geval worden de parameters per afzonderlijke vergelijking geschat, terwijl dit in het tweede geval voor alle modelparameters tegelijkertijd gebeurt. Beide kategorieën hebben hun voor en tegen. Schatting per afzonderlijke vergelijking vergt geen ingewikkelde algoritmen, maar is niet mogelijk in geval van "contemporaneous" kovarianties tussen storingstermen. In dat geval dient een "full-information" techniek worden toegepast. Een nadeel van deze laatste is dat een specifikatiefout in één der vergelijkingen reperkussies heeft voor het gehele model. Andere problemen welke ik in hoofdstuk 4 nader zal uitwerken, liggen op het vlak van geldigheid van operationalisering en theoriekonstruktie.

Voor een behandeling van een groot aantal voorhanden zijnde schattingstechnieken wordt verwezen naar met name ekonometrische en psychometrische literatuur. In hoofdstuk 3 worden enkele technieken behandeld welke speciaal van belang zijn voor de analyse van modellen met ongemeten variabelen en het daaruit voortvloeiende probleem van overidentifikatie. Van de "single-equation" technieken wordt behandeld de methode van instrumentele variabelen (IV). Deze is van belang wegens toepasbaarheid in het geval van random meetfouten (hoofdstuk 5) alsook vanwege het feit dat pad-analyse aantoonbaar een IV-techniek is (paragraaf 3 van dit hoofdstuk). Eveneens uit deze kategorie wordt behandeld de "two stage least squares" methode (2SLS). De relevantie hiervan is gelegen in het feit dat hier een oplossing wordt gevonden voor het probleem van overidentifikatie. Bovendien biedt deze techniek in sommige gevallen een oplossing voor het probleem van samenhang tussen regressor en storingsterm, welke zich zoals bleek vaak voordoet als gevolg van meetfouten. Sociologische en politikologische toepassingen zijn te vinden in Mason and Halter 1968, Duncan, Haller and Portes 1968 en van de Geer 1971: 217-220. Opmerkelijk is dat laatstgenoemde de 2SLS-methode geheel plaatst in het kader van rekursieve modellen. Dit is verwarrend aangezien deze techniek geenszins is gebonden aan rekursiviteit. Integendeel, meestal wordt zij juist toegepast bij niet-rekursieve modellen.

Van de "full - information" technieken is verreweg de belangrijkste de kategorie die valt onder de naam "full information maximum - likelyhood" (FIML). Zoals in de eerste paragraaf van dit hoofdstuk reeds gezegd, doen "contemporaneous" kovarianties tussen storingstermen zich veelal voor als gevolg van meetfouten, waarmee het belang van deze schattingsmethode voor deze studie is gegeven. Zij doet de laatste paar jaar veel opgang in de sociale wetenschappen. Zo steunen de door Jöreskog ontwikkelde analyse-modellen op deze methode. Een korte beschrijving van deze modellen volgt eveneens in hoofdstuk 3.

3. REKURSIEVE MODELLEN

In de sociologie en politikologie werden tot op heden vrijwel steeds rekursieve modellen gehanteerd. Een rekursief model is een struktureel model met uitsluitend asymmetrische relaties. Het is als zodanig een speciaal geval van de kategorie van strukturele vergelijkingen. Naast het feit dat men in de genoemde disciplines vaak te maken heeft met asymmetrische relaties, heeft aan deze praktijk bijgedragen de mathematische eenvoud van rekursieve modellen. Zo zijn er doorgaans minder schattingsproblemen dan in niet-rekursieve modellen.

Belangrijk voor de ontwikkeling van rekursieve modellen was de ekonometrist Wold. Hij bestudeerde de mogelijkheid van regressieanalyse als instrument van kausale analyse en toonde aan dat rekursieve systemen van regressievergelijkingen kunnen worden behandeld met OLS, uitgevoerd op de afzonderlijke vergelijkingen. (Wold, 1954, 1964, 1969). Deze ontwikkeling stond waarschijnlijk geheel los van eerdere pogingen tot hantering van regressie-analyse in het kausale onderzoek, namelijk die van de bioloog en genetikus Sewall Wright. Zijn methode, die hij pad-analyse noemde, gebruikte hij voornamelijk voor het toetsen van Mendeliaanse wetten (Wright, 1921, 1934, 1960). Aanvankelijk werden uit bekende parameters in deze wetten korrelaties afgeleid, maar in latere toepassingen van de techniek vindt een omgekeerde procedure plaats en worden modelparameters geschat op basis van empirisch gekonstateerde korrelaties.

Op een enkele uitzondering na (Tukey 1954, Wold 1956, 1969, Moran 1961, Hauser en Goldberger 1971), is Wright's techniek niet onder de aandacht van ekonometristen geweest. Dit lijkt merkwaardig maar de reden hiervan zou kunnen zijn de zwakke statistische status van deze techniek, waardoor zij mogelijk weinig aantrekkingskracht op de ekonometristen heeft uitgeoefend. Zo vindt men bij Wright niet een volledige opsomming van alle te stellen assumpties voor het schatten van modelparameters (padkoëfficiënten), worden steekproefkorrelaties onvoldoende gescheiden van populatiekorrelaties en wordt door volgelingen van Wright een statistisch niet houdbare ad hoc oplossing gesuggereerd voor het probleem van overidentifikatie. Verder kan hieraan ook hebben bijgedragen het feit dat Wright, hoewel niet strikt noodzakelijk, de pad-analyse verrichtte op basis van een korrelatie- in plaats van een in de ekonometrie gebruikelijke kovariantie-matrix. Een laatste oorzaak kan zijn gelegen in de problemen die ontstaan bij de analyse van niet-rekursieve modellen. Op hun beurt zijn sociologen zich op enkele uitzonderingen na (zie o.a. Blalock 1961, Duncan, Haller en Portes 1968) niet bewust dat voor problemen welke zich voordoen bij pad-analyse oplossingen zijn te vinden in de ekonometrische en psychometrische literatuur. In hoofdstuk 3 wordt dit tekort aangetoond en worden tevens enkele van de oplossingen geschetst.

rekursiviteit

De omschrijving van een rekursief model als een struktureel model met asymmetrische relaties behoeft precisering. De afspraak is dat een endogene variabele met een lager subskript een oorzaak is van een endogene variabele met een hoger subskript. Een model nu is rekursief indien een ordening bestaat in de variabelen alsook in de vergelijkingen, zodanig dat in de i-de vergelijking de waarde wordt bepaald van endogene variabele y_i als een funktie van exogene variabelen en endogene variabelen y_j $i > j$ welke als regressoren respektievelijk endogene regressoren in de i-de vergelijking aanwezig zijn. Ook deze omschrijving wordt aanstonds nog toegespitst naar aanleiding van te stellen assumpties.

Het uitgangspunt is de strukturele vorm in (2.9)

$$\underset{\sim}{B}\underset{\sim}{y}_n + \underset{\sim}{\Gamma}\ \underset{\sim}{x}_n = \underset{\sim}{u}_n \quad n,\ldots,N$$

met de voor strukturele modellen gebruikelijke assumpties. Voor de matrix $\underset{\sim}{B}$ $(=\beta_{ij})$ geldt de afspraak dat het eerste subskript slaat op een kausaal afhankelijke variabele en het tweede subkript op een kausaal onafhankelijke variabele. Dit impliceert voor rekursieve modellen

$$\beta_{ij} = 0 \text{ voor } i < j$$

De assumptie van rekursiviteit houdt dus in dat de bovendriehoek van $\underset{\sim}{B}$ uitsluitend elementen gelijk aan nul bevat. Kortweg duid ik deze assumptie aan als "$\underset{\sim}{B}$ driehoekig". Dit is de belangrijkste assumptie voor rekursieve modellen. Voor Goldberger is deze assumptie niet alleen een noodzakelijke maar tevens een voldoende voorwaarde om te spreken van rekursiviteit. (Goldberger 1964: 354). Laatstgenoemde onderscheidt daarnaast een zogenaamd diagonaal rekursief model. Dit is een rekursief model waarvoor de assumptie geldt van afwezigheid van "contemporaneous" kovarianties tussen storingstermen. Deze assumptie noem ik kortweg "$\underset{\sim}{\Sigma}$ diagonaal", waar $\underset{\sim}{\Sigma}$ is gedefinieerd als in (2.11). Bezien we thans opnieuw formule (2.13) betreffende kovarianties tussen endogene variabelen en storingstermen

$$(2.14) \qquad \underset{\sim}{\Sigma}_{yu} = \sigma(y_i u_j) = \underset{\sim}{B}^{-1} \underset{\sim}{\Sigma}$$

Aangezien voor $\underset{\sim}{B}^{-1}$ dezelfde nulspecifikaties (eksklusie-restrikties) gelden als voor $\underset{\sim}{B}$ en bovendien geldt "$\underset{\sim}{\Sigma}$ diagonaal", gelden voor $\underset{\sim}{\Sigma}_{yu}$ dezelfde nulspecifikaties als voor $\underset{\sim}{B}$.

$$(2.15) \qquad \sigma(y_i u_j) = 0 \quad \text{voor } i < j$$

De endogene variabelen y_i kovariëren niet met storingstermen u_j voor $i < j$. Voorts kunnen in een bepaalde vergelijking j uitsluitend variabelen y_i $i < j$ optreden als endogene regressoren. Op grond hiervan volgt uit (2.15) dat in rekursieve modellen de endogene regressoren in een bepaalde vergelijking ongekorreleerd zijn met de storingsterm uit die vergelijking. Uiteraard mits voldaan is aan de gestelde assumpties "$\underset{\sim}{B}$ driehoekig" en "$\underset{\sim}{\Sigma}$ diagonaal". Indien beide laatstgenoemde assumpties worden gemaakt, zal ik voortaan spreken van een rekursief model. De gestelde assumpties noem ik *model-assumpties* waarmee zijn bedoeld assumpties ten aanzien van de specifikatie van het model. Deze worden gebaseerd op theoretische en of a-priori inzichten. Daarnaast dienen nog andere assumpties te worden gemaakt, zoals die welke zijn genoemd in paragraaf 2 van hoofdstuk 2. Deze worden voortaan *systeem-assumpties* genoemd. Hieronder valt ook de assumptie van afwezigheid van meetfouten. Kenmerkend voor systeem-assumpties is dat zij nodig zijn voor de technische uitvoering van de analyse.

Op grond van het bovenstaande kunnen een tweetal konklusies worden getrokken: (a) endogene regressoren kunnen t.a.v. vergelijkingen waarin zij participeren worden beschouwd als exogene variabelen (zie (2.15)) en (b) de vergelijkingen vormen niet langer een interdependent systeem ($\underset{\sim}{\Sigma}$ diagonaal). Hierop kan een nieuwe konklusie worden gebaseerd welke inhoudt, dat met het oog op schatting van strukturele parameters alle endogene variabelen in (2.11) naar het rechterdeel van de respektievelijke vergelijkingen kunnen worden overgebracht, om vervolgens per afzonderlijke vergelijking OLS toe te passen. Strikt genomen is nog één model-assumptie nodig wil er geen sprake zijn van interdependentie in de vergelijkingen. Deze betreft de aanname dat geen "cross equation" restrikties aan het model zijn opgelegd. Bedoeld zijn restrikties die te maken hebben met de verhouding van strukturele parameters uit verschillende vergelijkingen ten opzichte van elkaar. Is aan alle drie genoemde model-assumpties voldaan dan kan de analyse van een rekursief model inderdaad worden teruggebracht tot een serie eenvoudige "single equation" regressie-analyses. Elke vergelijking behelst de konditionele verdeling van één variabele gegeven de waarden van een stel onafhankelijke variabelen, inklusief endogene regressoren. Bekend is dat in zulke gevallen de verkregen OLS schatters asymptotisch gelijk zijn aan ML schatters, mits mag worden uitgegaan van de normale systeem-assumpties uit de regressie-analyse, aangevuld met de assumptie van normaal verdeelde storingstermen (Wonnacott and Wonnacott: 1970: 34 en 196, Christ 1966:

395). Land komt op enigszins omslachtige wijze tot eenzelfde bevinding (Land 1973).

Het bovenstaande geldt uiteraard uitsluitend indien is voldaan aan de gestelde assumpties. Een voor de verdere hoofdstukken belangrijk geval waarin de assumpties worden verstoord, doet zich voor indien een met meetfout gemeten variabele in meerdere vergelijkingen als regressor optreedt. Zoals eerder gezegd, leidt dit tot "contemporaneous" kovarianties tussen storingstermen, zodat niet langer kan worden volstaan met OLS.

In het onderstaande wordt een tweetal typen rekursieve modellen behandeld, te weten het pad-analyse model en het model van faktor-analyse. Beide vormen van analyse worden zeer veel toegepast in de sociologie en politikologie, hoewel faktor-analyse vrijwel steeds wordt verricht met het oog op data-reduktie en niet zozeer voor het zoeken naar kausale strukturen achter korrelatiematrices. Een derde analysevorm welke thuishoort in de kategorie van rekursieve modellen, te weten de Simon-Blalock analyse, wordt voldoende bekend verondersteld.

pad-analyse

Bij de formulering van zijn padmodel gebruikte Wright een notatiewijze welke afwijkt van die welke gangbaar is in de ekonometrie. Hoewel ik aan deze laatste de voorkeur geef, wordt Wright's model hieronder in zijn eigen notatiewijze behandeld, zij het omgezet in matrixnotatie.

Uitgangspunt van Wright is een rekursief stelsel vergelijkingen, één voor elke expliciete variabele uit het model. Alle variabelen met een koëfficiënt ongelijk aan één staan in het rechterdeel van de vergelijkingen en alle storingstermen krijgen een parameter toegewezen. In matrix-notatie

$$(2.16) \qquad \underset{\sim}{z}_n = \underset{\sim}{A}\underset{\sim}{z}_n + \underset{\sim}{D}\underset{\sim}{u}_n \qquad n=1,\dots,N$$

$$\underset{\sim}{A} = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{L1} & \cdots & a_{L(L-1)} & 0 \end{pmatrix}, \quad \underset{\sim}{z}_n = \begin{pmatrix} z_{1n} \\ \vdots \\ z_{Ln} \end{pmatrix}, \underset{\sim}{u}_n = \begin{pmatrix} u_{1n} \\ \vdots \\ u_{Ln} \end{pmatrix}$$

$\underset{\sim}{A}$ is een $L*L$ matrix met strukturele parameters (padkoëfficiënten), $\underset{\sim}{z}_n$ is een $L*1$ vektor met expliciete variabelen, $\underset{\sim}{u}_n$ is een $L*1$ vektor met storingstermen en $\underset{\sim}{D}$ is een $L*L$ diagonale matrix met parameters behorende bij de storingstermen. Uit de struktuur van $\underset{\sim}{A}$ blijkt welke variabelen exogeen zijn. Bevat de i-de rij van $\underset{\sim}{A}$ uitsluitend nullen dan is de variabele met suffix i exogeen. Voorts is kenmerkend dat de variabelen zijn gegeven in standaardskores.[3)] Duncan (1966) stelde een formule op, door hem het basistheorema van de pad-analyse genoemd, waarmee korrelaties kunnen worden geschreven in termen van strukturele parameters en korrelatiekoëfficiënten.

$$(2.17) \qquad \rho_{ij} = \sum_q a_{iq}\,\rho_{jq}$$

Het symbool ρ staat voor een populatiekorrelatie[4)] terwijl i steeds aangeeft een endogene variabele en j slaat op hetzij een exogene hetzij een endogene variabele. Voor q geldt hetzelfde als voor j, evenwel met de beperking dat het hier uitsluitend variabelen betreft met een *direkt* kausaal effekt op variabele i. Het oogmerkt van Duncan met formule (2.17) was het interpreteren van kausale modellen. Door een herhaalde toepassing van deze formule kunnen korrelaties worden uiteengelegd in kausale paden. Dat wil zeggen, de formule kan worden toegepast voor ρ_{jq} in het rechterdeel van (2.17) en wel voor elke q. Wanneer dit proces wordt herhaald totdat j en q beide exogene variabelen zijn, wordt een overzicht verkregen van direkte en indirekte effekten tussen de variabelen i en j (zie Duncan 1966). Formule (2.17) kan óók worden gebruikt voor het verkrijgen van zogenaamde padschattingsvergelijkingen. Deze vergelijkingen zijn identiek aan de normaalvergelijkingen uit de OLS-procedure, evenwel met dit verschil dat in het eerste geval wordt uitgegaan van gestandaardiseerde skores, terwijl bij OLS gewoonlijk ruwe skores worden gehanteerd. Daardoor zijn deze vergelijkingen niet opgebouwd uit ongestandaardiseerde parameters en kovarianties zoals in het geval van OLS, maar uit gestandaardiseerde parameters (padkoëfficiënten) en korrelaties.

Deze toepassing van formule (2.17) behoeft enige toelichting. Zoals bekend kan een stel normaalvergelijkingen worden verkregen via minimalisering van de regressie-residuen, dit laatste met behulp van differentiaalrekening. Een andere methode welke onder normale omstandigheden toepasbaar is, is die van doorvermenigvuldiging (zie Christ 1966: 404-407, Wonnacott 1970: 149-151 en Hays 1963: 567-569). Hierbij wordt zowel het linker- als het rechterdeel van een vergelijking achtereenvolgens vermenigvuldigd met alle onafhankelijke variabelen uit de vergelijking (N.B. dit zijn alle variabelen met een direkte invloed op de afhankelijke variabele in het linkerdeel van de vergelijking.) Neemt men vervolgens verwachtingswaarden van de ontstane produkten, dan verkrijgt men een stel normaalvergelijkingen. Dit stel heeft steeds precies evenveel vergelijkingen als onbekenden, zodat de parameters te bepalen zijn. Deze procedure toegepast op (2.16) geeft het volgende beeld. Postmultiplikatie van (2.16) met $\underset{\sim}{z}'_n$ geeft na het nemen van verwachtingswaarden

$$E(\underset{\sim}{z}_n \underset{\sim}{z}'_n) = E(\underset{\sim}{A} \underset{\sim}{z}_n \underset{\sim}{z}'_n) + E(\underset{\sim}{D} \underset{\sim}{u}_n \underset{\sim}{z}'_n)$$

ofwel

$$(2.18) \qquad \underset{\sim}{\Sigma}_{zz} = \underset{\sim}{A}\,\underset{\sim}{\Sigma}_{zz} + \underset{\sim}{D}\,\underset{\sim}{\Sigma}_{uz}$$

waar $\underset{\sim}{\Sigma}_{zz}$ $[=\sigma(z_i z_j)]$ en $\underset{\sim}{\Sigma}_{uz}[=\sigma(u_i z_j)]$ $L*L$ matrices zijn met korrelaties tussen expliciete variabelen onderling en tussen impliciete- en expliciete variabelen. Omdat de vergelijking voor afhankelijke variabele z_i uitsluitend wordt doorvermenigvuldigd met variabelen z_j $j < i$, kan de aandacht beperkt worden tot de onderdriehoeken van $\underset{\sim}{\Sigma}_{zz}$ en $\underset{\sim}{\Sigma}_{uz}$. De matrix $\underset{\sim}{\Sigma}_{uz}$ verdient enige aandacht. Stel dat de vektor $\underset{\sim}{z}_n$ uitsluitend endogene variabelen bevat. In dat geval vertoont $\underset{\sim}{\Sigma}_{uz}$ gelijkenis met $\underset{\sim}{\Sigma}_{yu}$ in (2.13). Het enige verschil is dat $\underset{\sim}{\Sigma}_{uz}$ als gevolg van postmultiplikatie de getransponeerde is van $\underset{\sim}{\Sigma}_{yu}$. Dit betekent dat in het gestelde geval niet de bovendriehoek maar de onderdriehoek van $\underset{\sim}{\Sigma}_{uz}$ uitsluitend nullen bevat. Van exogene variabelen weten we dat zij per definitie ongekorreleerd zijn met storingstermen, zodat het gestelde ten aanzien van $\underset{\sim}{\Sigma}_{uz}$ ongewijzigd blijft indien $\underset{\sim}{z}_n$ ook exogene variabelen bevat. Gezien dit resultaat kan van $\underset{\sim}{D}\,\underset{\sim}{\Sigma}_{uz}$ in (2.18) worden gezegd dat de onderdriehoek uitsluitend nullen bevat. Gebruikmakend van superskript Δ om aan te geven dat het gaat om onderdriehoeken, kan naar aanleiding van het bovenstaande en op basis van (2.18) worden geschreven

$$(2.19) \qquad \underset{\sim}{\Sigma}^{\Delta}_{zz} = \underset{\sim}{A}\,\underset{\sim}{\Sigma}^{\Delta}_{zz}$$

Deze vergelijking bestaat uit $\frac{1}{2}L(L-1)$ padschattingsvergelijkingen met evenveel padkoëfficiënten. Analoog als deze procedure is aan OLS, levert dit als konklusie op dat padkoëfficiënten identiek zijn aan gestandaardiseerde partiële regressiekoëfficiënten. Het valt gemakkelijk in te zien dat exakt dezelfde resultaten als in (2.19) worden verkregen wanneer formule (2.17) wordt toegepast op elke afhankelijke variabele x_i in het model.

Zodra sprake is van onvolledige rekursiviteit treden komplikaties op. Als uitgangspunt neem ik de i-de vergelijking uit (2.16). De verzameling onafhankelijke variabelen in deze vergelijking duid ik aan met Q_i. Een variabele in deze verzameling is hetzij een exogene variabele hetzij een endogene regressor. Bewezen werd dat voor beide geldt dat er geen samenhang is met de storingsterm u_i. Nu wordt er door een aantal pad-analytici terecht op gewezen dat de variabelen in Q_i niet de enige variabelen zijn waarvoor dit geldt. Alle variabelen in het linkerdeel van voorafgaande vergelijkingen hebben subskript j j < i en vertonen om die reden dezelfde eigenschap. De verzameling van deze variabelen wordt aangeduid met P_i. In het speciale geval van volledige rekursiviteit zoals hierboven het geval was, geldt uiteraard $Q_i=P_i$. In een onvolledig rekursief model is Q_i evenwel een deelverzameling van P_i. Door sommige pad-analytici nu wordt voorgesteld alle variabelen in P_i te betrekken in de schattingsprocedure. Dit leidt tot meerdere schattingen van een en dezelfde strukturele parameter, welke schattingen normaliter niet konvergeren. De vraag rijst wat in een dergelijke situatie te doen. Boudon stelde voor een kleinste kwadraten oplossing te berekenen voor de verkregen alternatieve schattingen (Boudon 1971: 213-215). Goldberger bewees dat dit echter leidt tot inefficiënte schatters (Goldberger 1968). Zijn voorstel is om uitsluitend de verzameling Q_i te hanteren in de schattingsprocedure, met andere woorden om OLS toe te passen.

Een tweede mogelijkheid welke door pad-analytici wordt gesuggereerd, is om alle strukturele parameters a priori ongelijk aan nul te specificeren, met ander woorden om uit te gaan van een volledig rekursief model, óók al zou de theorie wijzen op onvolledige rekursiviteit. Pas indien mocht blijken dat de bewuste parameters inderdaad gelijk aan nul dan wel verwaarloosbaar klein blijken te zijn, worden zij alsnog gelijkgesteld aan nul. Vervolgens worden in dit geherspecificeerde model opnieuw de schattingen verricht door toepassing van OLS op de afzonderlijke vergelijkingen. Deze werkwijze wordt in feite voorgesteld door Duncan (zie Duncan 1966: 8 en 9). Enige wenken in deze richting geeft ook van de Geer (1971: 21).

Ik heb bezwaar tegen deze werkwijze, aangezien zij een inkonsekwentie bevat. Nooit immers wordt in rekursieve modellen nagegaan of de parameters in de bovendriehoek van $\underset{\sim}{A}$ (gevolg van de hypothese van asymmetrie) terecht gelijk gesteld zijn aan nul. Hier vertrouwt men kennelijk volledig op de theorie en of a priori inzichten. Waarom dit dan ook niet gedaan ten aanzien van eksklusie-restrikties in de onderdriehoek van $\underset{\sim}{A}$? De aangewezen oplossing is in dit geval een modelspecifikatie konform theoretische en à priori inzichten, om vervolgens de schatting per afzonderlijke vergelijking met OLS af te handelen.

Hoewel hier wordt voorgesteld geen variabelen buiten de verzameling Q_i in de schattingsprocedure te betrekken, maak ik een uitzondering voor modellen met ongemeten variabelen. In het volgende hoofdstuk zal blijken dat pad-analyse toepasbaar is op modellen met ongemeten variabelen, door variabelen welke in P_i doch *niet* in Q_i aanwezig zijn te hanteren als instrumentele variabelen.

faktor-analyse

Interessant is de vraag of een faktormodel kan worden gezien als een struktureel model. Uitgangspunt is het volgende faktormodel

$$(2.20) \qquad \underset{\sim}{y}_n = \underset{\sim}{\Lambda}\, \underset{\sim}{f}_n + \underset{\sim}{u}_n \qquad n=1,\ldots,N$$

$\underset{\sim}{y}_n$ is een $G*1$ vektor met inputvariabelen, $\underset{\sim}{\Lambda}$ is een $G*H$

matrix met faktorladingen, $\underset{\sim}{f}_n$ is een $H * 1$ vektor met faktoren en $\underset{\sim}{u}_n$ is een $G * 1$ vektor met uniciteiten. In faktormodellen worden over het algemeen de volgende assumpties gemaakt

(2.21) $E(\underset{\sim}{f}_n \underset{\sim}{u}_n') = \underset{\sim}{0}_n$ en $E(\underset{\sim}{u}_n \underset{\sim}{u}_n') = \underset{\sim}{\Psi}$ (diagonaal)

Met andere woorden, de assumpties zijn dat er geen samenhang is tussen uniciteiten en faktoren noch tussen uniciteiten onderling. Hantering van een faktormodel als struktureel model brengt een wijzing in terminologie met zich mee. Inputvariabelen heten endogene variabelen, faktorladingen staan voor modelparameters, faktoren staan voor exogene ongemeten variabelen en uniciteiten heten storingstermen. (In hoofdstuk 4 wordt aan de betekenis van uniciteiten een uitbreiding gegeven). De matrix $\underset{\sim}{\Psi}$ wordt in een struktureel model vervangen door $\underset{\sim}{\Sigma}$. Uit (2.21) blijkt dat voor storingstermen ($\underset{\sim}{u}_n$) in een faktormodel dezelfde assumpties gelden als in een struktureel model. Opvallend is voorts dat geen parameters zijn toegevoegd aan de inputvariabelen. [Zie (2.20)]. Dit betekent dat geen strukturele relaties mogen voorkomen tussen endogene variabelen.

Nog altijd is geen antwoord gegeven op de vraag of een faktormodel een kausale interpretatie toestaat. Deze vraag is tweeledig: (a) kunnen de relaties in een faktormodel worden geïnterpreteerd als kausale relaties en (b) indien ja, gaat het dan om symmetrische dan wel asymmetrische relaties? Ten aanzien van de eerste vraag kan worden gesteld dat faktor-analyse formeel kan worden beschouwd als een regressie-analyse met ongemeten regressoren. (Harman 1979:16,van de Geer 1971: 87). Een afdoende argumentatie voor regressie-analyse als kausaal-analytisch instrument werd gegeven in hoofdstuk 1, waarmee deze vraag voldoende is beantwoord. Ten aanzien van de tweede vraag wordt gewezen op de bevinding in hoofdstuk 2 dat een noodzakelijke voorwaarde voor de assumptie van afwezigheid van samenhang tussen storingstermen en onafhankelijke variabelen, hier de afwezigheid van symmetrische relaties is. Aangezien de systeem-assumptie " $\underset{\sim}{\psi}$ diagonaal" bij faktor-analyse steeds dient worden gemaakt, moet de konklusie zijn dat faktor-analyse uitsluitend toepasbaar is op rekursieve modellen.

Er zijn evenwel vele beperkingen verbonden aan hantering van faktor-analyse als struktureel model. Zo kunnen geen model-assumpties worden gemaakt, mogen endogene variabelen niet kausaal verbonden zijn en kunnen geen indirekte relaties in het model voorkomen. Een handicap is voorts dat faktorkorrelaties in modellen met drie of meer faktoren niet kausaal kunnen worden geïnterpreteerd. (De draagwijdte van deze opmerking zal eerst goed in het volgende hoofdstuk duidelijk worden.) In latere uitbreidingen van faktormodellen zoals "second order" faktor-analyse, Jöreskog's algemene kovariantiemodel en diens LISREL-programma, is zulks mogelijk. Verschillende van de andere genoemde bezwaren worden ondervangen in de eveneens door Jöreskog ontwikkelde "confirmatory" faktor-analyse. Hierop zal in hoofdstuk 3 worden teruggekomen.

3. MODELLEN MET ONGEMETEN VARIABELEN

Bij een voortschrijdende theorievorming neemt ook het gebruik van abstrakte begrippen toe. Als gevolg daarvan geraakt de methodiek van strukturele vergelijkingen, in haar toepassing van respresentatie van theorieën, verweven met ingewikkelde meetproblemen. De meest toegepaste aanpak van deze problematiek is in dit kader een meervoudige operationalisering. Per abstrakt begrip worden meerdere indikatoren gekozen waarna men de keus heeft uit (a) de indikatoren samenvoegen tot een index en een strukturele analyse verrichten als betrof het een model met uitsluitend gemeten variabelen (indices) of (b) de strukturele analyse uitvoeren op basis van afzonderlijke indikatoren. De eerstgenoemde werkwijze wordt traditioneel gevolgd door sociologen en politikologen. Aan het begin van de jaren zestig onderkende men echter voordelen aan strukturele analyse rechtstreeks op basis van indikatoren. Men zie bijvoorbeeld Curtis en Jackson 1962, Siegel en Hodge 1968, Costner 1969, Alwin 1974. Hun voornaamste argumenten betreffen de externe validering van indikatoren, de identifikatie van strukturele parameters en de opsporing van spurieuze relaties.

Tegen beide werkwijzen kunnen bezwaren worden ingebracht. Een nadeel van de traditionele aanpak is dat impliciet wordt aangenomen dat een index een perfekte representatie geeft van een ongemeten variabele. In hoofdstuk 4 en 5 zal blijken dat deze aanname berust op een fiktie. Ook de werkwijze sub b kan op bezwaren stuiten. In hoofdstuk 4 wordt aangetoond dat de validiteit van operationalisering en theorievorming kan worden aangetast, met name wanneer "full- information"-technieken worden gebruikt. De aan werkwijze sub b verbonden identifikatie- en schattingsproblemen worden hieronder uitgewerkt in de respektievelijke paragrafen 1 en 2. Procedures waarin verschillende van deze problemen worden ondervangen zijn het onderwerp van de paragrafen 2 en 3.

Alvorens hiertoe over te gaan, dient te kunnen worden beschikt over een begrippenkader. Kenmerkend voor de werkwijzen sub a en sub b is het onderscheiden van een deel in het strukturele model dat betrekking heeft op theoretisch relevante variabelen en een deel dat slaat op de variabelen via welke deze laatste worden waargenomen. Het eerste deel, de *hoofdtheorie*, bevat de gepostuleerde relaties tussen theoretisch relevante variabelen. Deze relaties noemt men strukturele relaties, welke worden gekwantificeerd met behulp van *strukturele parameters*. Het tweede deel, de *hulptheorie*, bevat de gemeten variabelen alsmede de relaties van deze laatste met de ongemeten variabelen uit de hoofdtheorie. Deze relaties noemt men epistemische relaties en in het verlengde hiervan spreek ik van *epistemische parameters*.

De vraag is wat precies moet worden verstaan onder een *ongemeten variabele*. Bedoeld is hier een theoretische variabele in de zin van een variabele los van de empirie. Het is een variabele ontdaan van problemen die optreden bij de operationalisering en waarneming. Het is per definitie een variabele zonder meetfout. Men zou ook kunnen zeggen dat het hier de zuivere vorm van een variabele betreft welke uitsluitend fungeert als denktechnisch hulpmiddel bij de behandeling van meetfouten.

Voorts wordt een tweetal kategorieën ongemeten variabelen onderscheiden, te weten *latente* variabelen en wat ik bij gebrek aan een betere term *errorvrije* variabelen zal noemen. Met latente variabelen worden bedoeld komplexe niet direkt meetbare variabelen van abstrakt en theoretisch niveau. Kenmerkend is dat zij meerdere betekenisaspekten hebben waardoor meervoudige operationalisering is gewenst. Bij hun meting treden meetfouten op welke naast de betrouwbaarheid vooral de geldigheid van meting betreffen. De observatie van latente variabelen geschiedt in termen van gemeten variabelen welke worden aangeduid als *indikatoren*. Met errorvrije variabelen bedoel ik variabelen van gering abstraktie-niveau bij de meting waarvan niettemin meetfouten zullen optreden. Deze betreffen vooral de betrouwbaarheid. Meervoudige operationalisering is niet nodig, tenzij deze wordt verricht in het kader van onderzoek naar de meetbetrouwbaarheid (zie hoofdstuk 5). De observatie van een errorvrije variabele noem ik meetresultaat of gemakshalve ook wel indikator. De begrippen errorvrije- en latente variabele worden in hoofdstuk 5 nader toegespitst.

1. IDENTIFIKATIE

Met het oog op een behandeling van identifikatie- en schattingsproblemen volgt eerst de formulering van een algemeen model met ongemeten variabelen. Over het algemeen wordt er in deze studie van uitgegaan dat alle variabelen in de hoofdtheorie van dit model van theoretisch nivo zijn, maar zoals zal blijken kunnen via eenvoudige specifikaties ook modellen met simpele direkt meetbare variabelen in de hoofdtheorie worden behandeld. Uit dit algemeen model kunnen modellen van de meest uiteenlopende vorm en komplexiteit worden afgeleid. Uitgangspunt is een model met G ongemeten endogene variabelen, verzameld in een $G*1$ vektor $\underset{\sim}{\eta}$ en H ongemeten exogene variabelen, verzameld in $H*1$ vektor $\underset{\sim}{\xi}$ Deze ongemeten variabelen en hun onderlinge relaties vormen gezamenlijk de hoofdtheorie. Strukturele relaties tussen ongemeten variabelen kunnen in de volgende verzameling strukturele vergelijkingen worden weergegeven.

(3.1) $$\underset{\sim}{B}\underset{\sim}{\eta} = \underset{\sim}{\Gamma}\underset{\sim}{\xi} + \underset{\sim}{\zeta} \qquad \text{(hoofdtheorie)}$$

$\underset{\sim}{B}$ is een $G*G$ en $\underset{\sim}{\Gamma}$ een $G*H$ matrix met strukturele parameters, terwijl $\underset{\sim}{\zeta}$ een $G*1$ vektor is met storingstermen. De variabelen $\underset{\sim}{\eta}$ en $\underset{\sim}{\xi}$ worden gemeten via respektievelijk P indikatoren y_i verzameld in een $P*1$ vektor $\underset{\sim}{y}$ en Q indikatoren x_i verzameld in een $Q*1$ vektor $\underset{\sim}{x}$. Ook hier wordt uitgegaan van afwijkingsskores. De indikatoren en hun epistemische relaties met de variabelen uit de hoofdtheorie vormen tesamen de hulptheorie, welke als volgt kan worden weergegeven

(3.2) $$\underset{\sim}{y} = \underset{\sim}{\Lambda}_y \eta + \underset{\sim}{\delta}$$

(hulptheorie)

(3.3) $$\underset{\sim}{x} = \underset{\sim}{\Lambda}_x \xi + \underset{\sim}{\varepsilon}$$

De indikatoren kunnen worden verzameld in een $(P+Q)*1$ vektor $\underset{\sim}{z}$

$$\underset{\sim}{z} = [\underset{\sim}{y}' \ \underset{\sim}{x}']'$$

De vektoren $\underset{\sim}{\delta}$ en $\underset{\sim}{\varepsilon}$ betreffen meetfouten in respektievelijk $\underset{\sim}{y}$ en $\underset{\sim}{x}$, terwijl $\underset{\sim}{\Lambda}_y$ en $\underset{\sim}{\Lambda}_x$ de P*G respektievelijk Q*H matrices zijn met epistemische parameters. Gezien het ontbreken van parameters welke de indikatoren onderling verbinden, wordt hier impliciet aangenomen dat er *geen direkte relaties* bestaan tussen indikatoren. Deze aanname vloeit voort uit de op goede gronden berustende hypothese dat de samenhang tussen twee indikatoren geheel en al tot stand komt via hun beider afhankelijkheid van de ongemeten variabele welke zij indiceren.

In dit algemeen model is niet-rekursiviteit toegestaan zowel in de hoofd- als in de hulptheorie. Bovendien kan zoals gezegd een model met variabelen zonder meetproblemen worden gespecificeerd. Zo wordt een model met uitsluitend meetfoutloze variabelen in de hoofdtheorie verkregen via de specifikaties $\underset{\sim}{\delta}=0$, $\underset{\sim}{\varepsilon}=0$, $\underset{\sim}{\Lambda}_x=\underset{\sim}{\Lambda}_y=\underset{\sim}{I}$, waar $\underset{\sim}{I}$ staat voor een eenheidsmatrix.Hoewel in de gegeven notatie wordt uitgegaan van het voor sociologen en politikologen verreweg meest belangrijke geval van indikatoren als *effekt*-variabelen, is een zodanige modelspecifikatie mogelijk dat deze fungeren als *oorzaak*-variabelen. Als gevolg van de specifikaties $\underset{\sim}{\varepsilon} = \underset{\sim}{0}$ en $\underset{\sim}{\Lambda}_x = \underset{\sim}{I}$ gaat het algemeen model over in een model waar H gemeten variabelen x_i fungeren als oorzaken van endogene variabelen η_i.

Wat betreft de hulptheorie gaat men in de praktijk van het sociologisch en politikologisch onderzoek meestal uit van *asymmetrische* epistemische relaties en wordt vrijwel steeds uitgegaan van het ontbreken van direkte relaties tussen indikatoren en ongemeten variabelen welke deze niet indiceren. Aldus vermindert het aantal te schatten parameters tot 2(P+Q), zijnde P+Q epistemische parameters en even zoveel meetfoutenvarianties. In een dergelijk model is overidentifikatie snel bereikt. Berekend kan worden dat, zelfs in het uit oogpunt van overidentifikatie meest gunstige geval van een volledig rekursieve hoofdtheorie met slechts één exogene variabele, men reeds met twee indikatoren per ongemeten variabele overidentifikatie verkrijgt.

In een overgeïdentificeerd model kunnen meerdere schattingen per parameter worden verkregen, welke meestal niet konvergeren. Het probleem is dan te komen tot één unieke schatter, waarbij men op twee manieren te werk kan gaan. Of het aantal schattingsvergelijkingen wordt gereduceerd tot het aantal dat juist voldoende is voor schatting òf elke vergelijking wordt betrokken in de schattingsprocedure, om vervolgens een gewogen gemiddelde te nemen van de alternatieve schattingen.[1] De meeste oplossingen werden tot op heden gezocht in de eerstgenoemde richting. Een van de meest simpele is een willekeurige keuze te doen uit de schattingsvergelijkingen, zoals onder andere werd voorgesteld door de ekonometrist Christ (1966: 408). Een nadeel hiervan is dat niet alle informatie wordt gebruikt en dat elke keuze leidt tot een andere verzameling schattingen. Een betere mogelijkheid is toepassing van principale komponenten analyse of "two stage least squares", waarover aanstonds meer. Ook technieken om alternatieve schattingen te verenigen tot één schatter werden veelvuldig beproefd. Een voorbeeld is de reeds genoemde en door Goldberger weerlegde kleinste kwadraten methode van Boudon. Een andere "oplossing", bestaande uit een ongewogen gemiddelde van alternatieve schattingen, voldoet evenmin als Boudon's methode aan het kriterium van efficiëntie. Het kan worden bewezen dat, wil men komen tot efficiënte schatters, dient te worden gewogen met de varianties en kovarianties van de alternatieve schatters. Hauser en Goldberger hierover: "An efficient estimation procedure should take into account the differences in the sampling variabilities of the conflicting estimates"

(Hauser and Goldberger 1971: 103). Genoemde auteurs laten zien dat dit bijvoorbeeld gebeurt bij de "generalized least squares" methode (zie ook Goldberger 1972, 1973b, 1974). In datzelfde artikel stellen zij dat andere schattingsmethoden waarin het probleem van overidentifikatie op bevredigende wijze wordt opgelost die zijn welke zijn gebaseerd op het "maximum likelyhood" principe (Hauser and Goldberger 1971: 99)

normering meeteenheid

Naast het bovengenoemde is er nog een identifikatieprobleem dat van een geheel andere orde is. Stel we hebben de volgende relatie tussen indikator x en latente variabele ξ: $x = \lambda \xi + \varepsilon$. Stel voorts dat ξ wordt vermenigvuldigd met faktor k. Deze manipulatie kan worden geneutraliseerd door λ te vermenigvuldigen met $\frac{1}{k}$. Immers: $x = (\lambda/k)\, k \xi + \varepsilon = \lambda \xi + \varepsilon$. Doordat de meeteenheid van ξ niet eenduidig kan worden vastgesteld, ontstaat een onbepaaldheid in λ. De algemene stelregel luidt dan ook, dat parameters niet zijn geïdentificeerd wanneer de meeteenheid (en dus de skorevektor) van variabelen waarmee deze parameters zijn verbonden, onbepaald is. Dit probleem van onderidentifikatie kan in tegenstelling tot het "normale" geval hierboven niet worden opgelost door toevoeging van indikatoren. Men dient gebruikt te maken van één van beide volgende mogelijkheden: (a) standaardisering van de ongemeten variabele bijvoorbeeld door te zorgen voor een gemiddelde van nul en een standaarddeviatie van één of (b) gelijkstelling van de epistemische parameter van één der indikatoren van ξ aan één. In dit laatste geval wordt in feite de meeteenheid van de latente variabele gelijk gesteld aan die van de indikator waarvan de parameter op één wordt gesteld.

Het verdient geen aanbeveling beide standaardiseringen *tegelijkertijd* toe te passen. Stel een rekursief model bevat in de hoofdtheorie de kausale keten $\xi_1 \longrightarrow \eta_1 \longrightarrow \eta_2$. Stel voorts dat ξ_1 en η_2 zonder meetfout zijn gemeten (ξ_1 = x_1 en $\eta_2 = y_2$) en dat y_1 het meetresultaat is van de errorvrije η_1. De bijbehorende vergelijkingen zijn

$$
(3.4) \qquad
\begin{aligned}
&\text{(a)} \quad \eta_1 = \gamma_{11} x_1 + \zeta_1 \\
&\text{(b)} \quad y_2 = \beta_{21} \eta_1 + \zeta_2 \\
&\text{(c)} \quad y_1 = \lambda_{y11} \eta_1 + \varepsilon_1
\end{aligned}
$$

Zoals gezegd ligt hier als standaardisering voor de hand $\lambda_{y11}=1$. Wordt nu bovendien $\sigma^2(\eta_1) = 1$ als standaardisering ingevoerd dan ontstaan problemen. Immers op basis van principes uit de pad-analyse zou blijken $\rho(y_1 y_2) = \rho(\eta_1 y_2) = \beta_{21}$. Met andere woorden, de dubbele standaardisering leidt tot de konklusie dat de errorvrije korrelatie $\rho(\eta_1 y_2)$ gelijk is aan de "geattenueerde" korrelatie $\rho(y_1 y_2)$. Deze konklusie is in strijd met wat de klassieke meettheorie ons leert (zie ook hoofdstuk 5). Hieruit kan als algemene stelregel worden afgeleid dat een dubbele standaardisering dient te worden vermeden.

2. SCHATTING

Verspreid in dit en het voorafgaande hoofdstuk kwam een aantal schattingproblemen naar voren welke het gevolg zijn van het opnemen van ongemeten variabelen in het model. Rekapitulerend kunnen deze problemen worden onderscheiden naar die welke optreden als gevolg van (a) errorvrije variabelen dan wel van (b) latente variabelen. Bij (a) betreft het ten eerste de samenhang tussen storingsterm en onafhankelijke variabele wanneer deze laatste met meetfout wordt gemeten en (vaak als gevolg hiervan) ten tweede "contemporaneous" kovarianties tussen meetfouten, welke *steeds* optreden wanneer een met meetfout gemeten variabele in meerdere vergelijkingen voorkomt. Aangaande (b) geldt het probleem van het niet voorhanden hebben van voor schatting benodigde skores alsook dat van overidentifikatie zodra de verhouding tussen het aantal indikatoren en dat van ongemeten variabelen een bepaalde limiet overschrijdt.

Genoemde schattingsproblemen kunnen in principe worden aangepakt met behulp van het drietal hieronder te bespreken technieken, te weten: instrumentele variabelen-, "two stage least squares"- en "full-information maximum-likelihood"-methoden. De problemen voortvloeiend (1) uit samenhang tussen storingsterm en onafhankelijke variabelen en (2) uit ontbrekende skores, laten zich in veel gevallen oplossen met behulp van een instrumentele variabelen methode. Een konsekwentie van deze aanpak, alsook van de aanwezigheid van meerdere indikatoren in modellen met latente variabelen, is in veel gevallen overidentificatie, waarvoor een oplossing mogelijk is middels toepassing van "two stage least squares". Het probleem van samenhangende storingstermen tenslotte laat zich behandelen met "full-information maximum likelihood"- schattingsmethoden.

Beide eerstgenoemde technieken komen aan de orde in het restant van deze paragraaf, waarbij zal blijken dat sterke raakvlakken bestaan tussen de methode van instrumentele variabelen, pad-analyse en faktor-analyse. In het kader van de "full-information maximum-likelihood"-methoden worden vervolgens in paragraaf 3 van dit hoofdstuk modellen behandeld welke werden ontwikkeld door Jöreskog.

instrumentele variabelen (IV)

Het probleem van samenhang tussen storingsterm en onafhankelijke variabelen, ontstaan door meetfouten in deze laatste, kan in principe worden opgelost middels een instrumentele variabele. De vektor met skores van zo'n variabele vervangt die van met meetfout gemeten variabelen tijdens de schattingsprocedure. Hoewel een rekursief model een specifiek geval is van een struktureel model en als zodanig minder algemeen, neem ik niettemin het eerstgenoemde als uitgangspunt voor een verduidelijking. Dit wordt gerechtvaardigd door het feit dat IV-procedures traditioneel worden toegepast bij schattingsproblemen welke voortvloeien uit niet-rekursiviteit, terwijl ik hieronder wil demonstreren hoe problemen voortvloeiend uit *meetfouten* met deze techniek kunnen worden opgelost. Aangezien een IV-procedure behoort tot de kategorie van "single-equation" technieken, wordt uitgegaan van één bepaalde vergelijking, welke wordt geacht thuis te horen in een rekursief model.

$$y = \underset{\sim}{\beta}' \underset{\sim}{\xi} + \zeta \tag{3.5}$$

Hier staat $\underset{\sim}{\xi}$ voor een vektor met errorvrije variabelen, $\underset{\sim}{\beta}$ staat voor een vektor met strukturele parameters terwijl ζ slaat op een storingsterm. In deze vergelijking is geen onderscheid gemaakt tussen endogene en exogene regressoren hetgeen wordt gerechtvaardigd door de aanname van rekursiviteit. Voorts is de regressant y *niet* geschreven als errorvrije variabele (η). Deze vereenvoudiging is toegestaan aangezien zoals bekend meetfouten in de afhankelijke variabele geen verstoring geven in de schatting van strukturele parameters. Voor $\underset{\sim}{\xi}$ geldt naar analogie van (3.3) de volgende meetvergelijking

$$\underset{\sim}{x} = \underset{\sim}{\lambda}_x \underset{\sim}{\xi} + \underset{\sim}{\varepsilon}$$

waar $\underset{\sim}{\varepsilon}$ aangeeft de meetfouten in $\underset{\sim}{x}$. Aangezien het gaat om *errorvrije* variabelen kan de standaardisering $\underset{\sim}{\lambda}_x = \underset{\sim}{I}$

worden ingevoerd, waar $\underset{\sim}{I}$ staat voor eenheidsvektor.

Uitgaande van een steekproef van N eenheden, waarvoor per vooronderstelling (3.5) geldt, gaat naar aanleiding van bovenstaande (3.5) over in

$$(3.6) \qquad \underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\zeta}^*$$

Hier staat $\underset{\sim}{y}$ voor de skorevektor van y, $\underset{\sim}{X}$ staat voor de skorematrix van $\underset{\sim}{x}$ terwijl $\underset{\sim}{\zeta}^*$ aangeeft de skorevektor van ζ^*. Voor deze laatste geldt $\underset{\sim}{\zeta}^* = \underset{\sim}{\zeta} - \underset{\sim}{\beta}'\underset{\sim}{\varepsilon}$ waaruit samenhang blijkt tussen $\underset{\sim}{X}$ en $\underset{\sim}{\zeta}^*$

$$\text{plim } 1/N \, (\underset{\sim}{X}'\underset{\sim}{\zeta}^*) \neq \underset{\sim}{0}$$

met als gevolg dat OLS leidt tot inkonsistente schatting (zie hoofdstuk 2 paragraaf 1). Konsistente schatting is echter mogelijk indien er variabelen $\underset{\sim}{z}$ te vinden zijn met skorematrix $\underset{\sim}{Z}$ welke *niet* samenhangen met $\underset{\sim}{\zeta}^*$ maar *wel* met $\underset{\sim}{X}$.

$$(3.7) \qquad \text{plim } (1/N \, \underset{\sim}{Z}'\underset{\sim}{\zeta}^*) = \underset{\sim}{0}, \quad \text{plim } (1/N \, \underset{\sim}{Z}'\underset{\sim}{X}) \neq \underset{\sim}{0}$$

Wordt bij OLS (3.6) gepremultipliceerd met $\underset{\sim}{X}'$, bij IV-schatting geschiedt dit met $\underset{\sim}{Z}'$. Na deling door N geeft dit

$$1/N \, \underset{\sim}{Z}'\underset{\sim}{y} = 1/N \, \underset{\sim}{Z}'\underset{\sim}{X}\underset{\sim}{\beta} + 1/N \, \underset{\sim}{Z}'\underset{\sim}{\zeta}^*$$

waaruit als IV-schatter van $\underset{\sim}{\beta}$ resulteert

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{Z}'\underset{\sim}{X})^{-1}\underset{\sim}{Z}'\underset{\sim}{y}$$

De vraag rijst welke kriteria moeten gelden bij de keuze van instrumentele variabelen. Bezien we daartoe de vereisten in (3.7). De eerstgenoemde behoeft thans geen toelichting meer. De tweede eis betreffende de korrelatie tussen een instrumentele variabele en de onafhankelijke variabele welke deze vervangt, hangt samen met de standaardfout van schatting voor $\hat{\underset{\sim}{\beta}}$. Deze is invers gerelateerd aan genoemde korrelatie (zie b.v. Johnston 1972: 280). Een derde te stellen voorwaarde welke nog niet werd genoemd, is dat een instrumentele variabele geen direkte invloed heeft op de afhankelijke variabele. Is die invloed er namelijk wél dan brengt invoering van elke instrumentele variabele een modelparameter, dat wil zeggen een onbekende, met zich mee waardoor een identifikatieprobleem ontstaat.

Zodra zich in een struktureel model overidentifikatie voordoet, is er meestal een duidelijke kategorie van variabelen die zich uitstekend lenen voor gebruik als instrumentele variabelen. Bedoeld zijn exogene variabelen welke niet in de te schatten vergelijking zijn opgenomen. Voor rekursieve modellen is er meestal zelfs een grotere kategorie welke wordt gevormd door de verzameling P_i op blz. 58. Bewezen werd dat variabelen in deze set niet samenhangen met de storingsterm uit de i-de vergelijking. Bovendien korreleren zij meestal in voldoende mate met de onafhankelijke variabelen aangezien zij er, behoudens uitzonderingsgevallen, kausaal mee verbonden zijn en hebben zij per definitie geen direkt effekt op de afhankelijke variabele y_i uit de i-de vergelijking.

Een toelichting op de geschetste procedure is mogelijk aan de hand van model (3.4). OLS-schatting van γ_{11} is zonder meer mogelijk. Niet echter die van β_{21} vanwege de samenhang van y_1 met de storingsterm welke ontstaat nadat in (3.4.b) y_1 is gesubstiteerd voor η_1. Echter x_1 voldoet als element van verzameling P_2 aan alle drie te stellen voorwaarden voor instrumentele variabelen, zodat op basis van een steekproef van N eenheden de volgende IV-schatting mogelijk is

$$\hat{\beta}_{21} = \sum_{n=1}^{N} x_{1n}y_{2n} / \sum_{n=1}^{N} x_{1n}y_{1n}$$

Strikt genomen moet als voorwaarde voor deze werkwijze de eis worden gesteld dat y_1 wordt gemeten in afwijkingsskores en *niet* in standaardskores (zoals gebruikelijk bij pad-analyse). Uit de klassieke meettheorie is namelijk bekend dat korrelaties worden verstoord door random meetfouten, hetgeen niet het geval is voor kovarianties. Van deze laatste eigenschap werd hierboven *impliciet* gebruik gemaakt welke mogelijkheid wegvalt indien wordt gestandaardiseerd.

Na bovenstaande uiteenzetting valt gemakkelijk in te zien dat eenzelfde werkwijze kan worden gevolgd in het geval de skores op een regressor geheel en al ontbreken. Een voorbeeld van een rekursief model met ontbrekende skores is het volgende, waarin één ongemeten variabele η_1, twee oorzaakvariabelen x_1 en x_2 en twee effektvariabelen y_2 en y_3. Deze vier laatste variabelen worden geacht *geen* indikatoren te zijn van η_1 maar als theoretisch relevante variabelen te zijn opgenomen in de hoofdtheorie. De strukturele vergelijkingen luiden hier

$$\text{(a)}\quad \eta_1 = \gamma_{11}x_1 + \gamma_{12}x_2 + \zeta_1$$

$$\text{(3.8)}\qquad \text{(b)}\quad y_2 = \beta_{21}\eta_1 + \zeta_2$$

$$\text{(c)}\quad Y_3 = \beta_{31}\eta_1 + \zeta_3$$

Vervolgens kan het volgende stelsel vergelijkingen worden uitgeschreven

$$\sigma(y_2x_1) = \beta_{21}\sigma(\eta_1x_1) = \beta_{21}\gamma_{11}\sigma^2(x_1) + \beta_{21}\gamma_{12}\sigma(x_1x_2)$$
$$\sigma(y_2x_2) = \beta_{21}\sigma(\eta_1x_2) = \beta_{21}\gamma_{11}\sigma(x_1x_2) + \beta_{21}\gamma_{12}\sigma^2(x_2)$$
$$\sigma(y_3x_1) = \beta_{31}\sigma(\eta_1x_1) = \beta_{31}\gamma_{11}\sigma^2(x_1) + \beta_{31}\gamma_{12}\sigma(x_1x_2)$$
$$\sigma(y_3x_2) = \beta_{31}\sigma(\eta_1x_2) = \beta_{31}\gamma_{11}\sigma(x_1x_2) + \beta_{31}\gamma_{12}\sigma^2(x_2)$$
$$\sigma(y_2y_3) = \beta_{31}\sigma(\eta_1y_2) = \beta_{31}\gamma_{11}\sigma(x_1y_2) + \beta_{31}\gamma_{12}\sigma(x_2y_2)$$

Op basis van deze vijf vergelijkingen kunnen de vier modelparameters worden geschat wanneer populatie(ko-)varianties worden vervangen door steekproefgrootheden. Het hierbij optredend probleem van overidentifikatie kan worden opgelost met "two stage least squares" (zie hierna).

In de beide bovenstaande gevallen is duidelijk herkenbaar een pad-analytische inslag, bestaande uit het doorvermenigvuldigen van een vergelijking niet alleen met de onafhankelijke variabelen uit die vergelijking, maar tevens met variabelen uit voorafgaande vergelijkingen. Deze laatste voldoen aan de eisen welke gelden voor instrumentele variabelen, zodat pad-analyse, toegepast op een model met ongemeten variabelen, is te kenschetsen als een instrumentele variabelen methode.

Via de hierboven ontwikkelde gedachtenlijn is het mogelijk de overeenkomsten te laten zien tussen IV, enerzijds en pad-analyse en faktor-analyse anderzijds. Als overgangsgebied tussen beide laatstgenoemde technieken wordt als voorbeeld een eenvoudig model met één latente variabele ξ en drie indikatoren x_i i=1,....,3 genomen waarvoor geldt (variabelen in afwijkingsskores)

$$x_1 = \lambda_1 \xi + \varepsilon_1, \quad x_2 = \lambda_2 \xi + \varepsilon_2, \quad x_3 = \lambda_3 \xi + \varepsilon_3$$

Pad-analytici zouden hier met het oog op de schatting van λ_1, λ_2 en λ_3 als volgt te werk gaan. De vergelijkingen worden paarsgewijs met elkaar vermenigvuldigd, met als resultaat

$$\begin{aligned}
\sigma(x_1 x_2) &= \lambda_1 \lambda_2 \, \sigma^2(\xi) \;;\; \sigma^2(x_1) = \lambda_1^2 \sigma^2(\xi) + \sigma^2(\varepsilon_1) \\
\sigma(x_1 x_3) &= \lambda_1 \lambda_3 \, \sigma^2(\xi) \;;\; \sigma^2(x_2) = \lambda_2^2 \sigma^2(\xi) + \sigma^2(\varepsilon_2) \\
\sigma(x_2 x_3) &= \lambda_2 \lambda_3 \, \sigma^2(\xi) \;;\; \sigma^2(x_3) = \lambda_3^2 \sigma^2(\xi) + \sigma^2(\varepsilon_3)
\end{aligned}$$

Er is sprake van onderidentifikatie gezien de zes vergelijkingen en zeven onbekenden, te weten: λ_1, λ_2, λ_3, $\sigma^2(\varepsilon_1)$, $\sigma^2(\varepsilon_2)$, $\sigma^2(\varepsilon_3)$ en $\sigma^2(\xi)$ terwijl bovendien de meetheid van ξ onbepaald is. Een normalisering als $\lambda_1=1$ lost beide identifikatieproblemen op. Voor λ_2 geldt dan bijvoorbeeld: $\lambda_2= \sigma(x_2x_3)/\sigma(x_1x_3)$, met als schatting van λ_2

$$\hat{\lambda}_2= \sum_{n=1}^{N} x_{2n}x_{3n}/\sum_{n=1}^{N} x_{1n}x_{3n}$$

We herkennen hierin een pad-analytische werkwijze. Toepassing van instrumentele variabelen nu levert exakt hetzelfde op. Substitutie van $\xi=x_1-\varepsilon_1$ in de vergelijking voor x_2 geeft

$$x_2= \lambda_2x_1 + \varepsilon_2^*$$

waar

$$\varepsilon_2^*= \varepsilon_2 - \lambda_2\xi_1$$

Als instrumentele variabele is bruikbaar x_3. Doorvermenigvuldiging en sommering geeft na simpele manipulatie

$$\hat{\lambda}_2=\sum_{n=1}^{N}x_{2n}x_{3n}/\sum_{n=1}^{N}x_{1n}x_{3n}$$

In het bovenstaande herkennen we tevens een principe uit de faktoranalyse, te weten schatting van modelparameters op basis van een kovariantiematrix betreffende inputvariabelen, welke laatste worden beschouwd als *afhankelijk* van ongemeten variabelen.

N.B. In hoofdstuk 2 werd aangetoond dat deze asymmetrische afhankelijkheid van inputvariabelen een noodzakelijke voorwaarde is voor de houdbaarheid van systeem-assumpties welke gelden voor faktor-analyse. Een en ander kan verder worden verduidelijkt aan de hand van het volgende faktor-model (zie(2.20)). (Variabelen in standaardskores).

(3.9) $$\underset{\sim}{y}_n= \underset{\sim}{\Lambda}\,\underset{\sim}{f}_n + \underset{\sim}{u}_n \qquad n=1,\ldots,N$$

Toepassing van OLS voor de schatting van modelparameter $\underset{\sim}{\Lambda}$ is onmogelijk, aangezien $\underset{\sim}{f}_n$ als ongemeten variabele ongeschikt is voor doorvermenigvuldiging. De vraag is of er instrumentele variabelen zijn. Bekijken we daartoe de matrix $\underset{\sim}{\Sigma}_{yu}$ met korrelaties tussen endogene variabelen en uniciteiten. Volgens (2.14) geldt $\underset{\sim}{\Sigma}_{yu} = \underset{\sim}{B}^{-1}\underset{\sim}{\Sigma}$. Aangezien in een faktormodel voor $\underset{\sim}{\Sigma}$ wordt geschreven $\underset{\sim}{\Psi}$ met overigens exakt identieke nulspecifikaties (zie 2.21) geldt

$$\underset{\sim}{\Sigma}_{yu} = \underset{\sim}{B}^{-1}\underset{\sim}{\Psi}$$

In een faktormodel geldt evenwel $\underset{\sim}{B} = \underset{\sim}{I}$ zodat

$$\underset{\sim}{\Sigma}_{yu} = \underset{\sim}{\Psi} \quad \text{(diagonaal)}$$

Er zijn geen samenhangen tussen inputvariabelen en unciteiten uit twee verschillende vergelijkingen. Dit betekent dat $\underset{\sim}{y}_n$ kan worden gehanteerd voor doorvermenigvuldiging. In termen van pad-analyse; de inputvariabelen $\underset{\sim}{y}_n$ behoren voor elke i-de vergelijking tot de verzameling P_i. Postmultiplikatie van (3.9) met $\underset{\sim}{y}'_n$ geeft na het nemen van verwachtingswaarden

$$E(\underset{\sim}{y}_n\underset{\sim}{y}'_n) = \underset{\sim}{\Lambda}\, E(\underset{\sim}{f}_n\underset{\sim}{y}'_n) + E(\underset{\sim}{u}_n\underset{\sim}{y}'_n)$$

ofwel

(3.10) $$\underset{\sim}{\Sigma}_{yy} = \underset{\sim}{\Lambda}\, E(\underset{\sim}{f}_n\underset{\sim}{y}'_n) + \underset{\sim}{\Sigma}_{yu}$$

waar $\underset{\sim}{\Sigma}_{yy}$ staat voor de korrelatiematrix van inputvariabelen. Zoals bekend is, geldt voor orthogonale faktormodellen dat faktorladingen kunnen worden opgevat als korrelatiekoëfficiënten behorende bij de relaties tussen inputvariabelen en faktoren, zodat geldt $E(\underset{\sim}{f}_n\underset{\sim}{y}'_n) = \underset{\sim}{\Lambda}'$. Op grond hiervan gaat (3.10) over in

(3.11) $$\underset{\sim}{\Sigma}_{yy} = \underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}$$

Tot zover de overeenkomsten met de IV-procedure.[2)] Een verschil treedt op bij de verdere afwikkeling, waarop hier niet nader wordt ingegaan. Voor meer informatie op dit punt raadplege men de betreffende literatuur. (Harman 1970, Lawley and Maxwell 1971).

Van de vier genoemde schattingsproblemen zijn er twee welke zoals bleek in principe met behulp van een IV-procedure kunnen worden aangepakt, te weten samenhang tussen onafhankellijke variabelen en storingstermen en het ontbreken van skores voor ongemeten variabelen. Van de twee overige kan overidentifikatie in sommige gevallen worden opgelost door toepassing van "two stage least squares". Een in dit opzicht nog bredere toepassingsmogelijkheid hebben "full - information maximum - likelihood" - technieken waarmee bovendien een oplossing kan worden gevonden voor het vierde probleem, te weten dat van "contemporaneous" kovariantie tussen storingstermen. Deze laatste komt aan de orde in paragraaf 3. Eerst volgt nu een uiteenzetting van de "two stage least squares"-methode.

two stage least squares (2SLS)

Hoewel de methode van "two stage least squares", evenals het geval is met de instrumentele variabelen methode, vooral wordt toegepast op problemen voortvloeiend uit niet-rekursiviteit, wordt hier de nadruk gelegd op een toepassing gericht op de meetfoutenproblematiek. Dit rechtvaardigt een behandeling gebaseerd op een rekursief model.

Vooral in rekursieve modellen is de kans groot dat meer instrumentele variabelen voorhanden zijn dan benodigd is voor schatting. Zijn voor niet-rekursieve modellen alle exogene variabelen geschikt als instrumentele variabelen, voor rekursieve modellen komen daar nog bij alle endogene variabelen uit "voorafgaande" vergelijkingen. Met elke instrumentele variabele kan een zogenaamde quasi-normaal-vergelijking (Christ 1966) worden gekreëerd. Hierdoor ontstaat veelal een stelsel vergelijkingen waarvan het aantal dat van de te schatten modelparameters overtreft. De vraag is hoe dit aantal zodanig kan worden teruggebracht dat evenveel vergelijkingen als onbekende ontstaan. Een mogelijkheid is dié instrumentele variabele te kiezen welke het sterkst korreleert met de te vervangen variabele.[3] Wellicht meer voor de hand liggend is echter een tweefasige regressie-analyse. In de eerste fase wordt een met meetfout gemeten onafhankelijke variabele x geregresseerd op *alle* variabelen uit het model die niet samenhangen met de storingsterm uit de te schatten vergelijking. (De hieruit resulterende variabele $\hat{x}$ kan worden beschouwd als een geschikte instrumentele variabele voor x.) Substitutie van $\hat{x}$ voor x geeft een vergelijking waarop vervolgens OLS kan worden toegepast.

Voor een verduidelijking ga ik uit van de vergelijking

$$(3.12) \qquad \underset{\sim}{y} = \underset{\sim}{\xi}\underset{\sim}{\beta} + \underset{\sim}{\zeta}$$

Stel dat een *deel* van de variabelen in $\underset{\sim}{\xi}$ met meetfouten is gemeten. Dit leidt tot partitionering van $\underset{\sim}{X}$, $\underset{\sim}{X} = [\underset{\sim}{X}_*\underset{\sim}{X}_\Delta]$, waar $\underset{\sim}{X}$ staat voor de *skore*-matrix van onafhankelijke variabelen in (3.12) en waar $\underset{\sim}{X}_*$ is de *skore*- matrix van variabelen welke mét en $\underset{\sim}{X}_\Delta$ welke zónder meetfouten zijn gemeten. Substitutie in (3.12) geeft

$$(3.13) \qquad \underset{\sim}{y} = [\underset{\sim}{X}_*\underset{\sim}{X}_\Delta]\ \underset{\sim}{\beta} + \underset{\sim}{\zeta}^*$$

waar $\underset{\sim}{\zeta}^*$, op analoge wijze als bij de IV-procedure hierboven, samenhangt met de matrix $[\underset{\sim}{X}_*\ \underset{\sim}{X}_\Delta]$. Een oplossing voor het aldus ontstane schattingsprobleem is $\underset{\sim}{X}_*$ te regresseren op alle variabelen in het model welke niet samenhangen met $\underset{\sim}{\zeta}^*$. Dit zijn vooreerst alle variabelen in verzameling P_i. De skorematrix voor deze variabelen duiden we evenals hierboven aan met $\underset{\sim}{Z}$. Bovendien komen in aanmerking de variabelen in $\underset{\sim}{X}_\Delta$. Kortom het gaat om de volgende theoretische regressievergelijking

$$\underset{\sim}{X}_* = \underset{\sim}{Z}_* \underset{\sim}{\Pi} + \underset{\sim}{V}$$

waar $\underset{\sim}{Z}_*$ is gedefiniëerd als $\underset{\sim}{Z}_* = [\underset{\sim}{Z}\ \underset{\sim}{X}_\Delta]$. De OLS schatter voor X_* is

$$\hat{\underset{\sim}{X}}_* = \underset{\sim}{Z}_* \hat{\underset{\sim}{\Pi}}$$

waar

$$\hat{\underset{\sim}{\Pi}} = (\underset{\sim}{Z}_*' \underset{\sim}{Z}_*)^{-1} \underset{\sim}{Z}_*' \underset{\sim}{X}_*$$

Substitutie van $\hat{\underset{\sim}{X}}_*$ in (3.13) geeft

$$\underset{\sim}{y} = [\hat{\underset{\sim}{X}}_* \underset{\sim}{X}_\Delta]\underset{\sim}{\beta} + \underset{\sim}{\zeta}^*$$

Op deze vergelijking kan vervolgens OLS worden toegepast, waarmee de volgende 2SLS-schatters worden verkregen.

$$\hat{\underset{\sim}{\beta}} = \{[\hat{\underset{\sim}{X}}_* \underset{\sim}{X}_\Delta]' [\hat{\underset{\sim}{X}}_* \underset{\sim}{X}_\Delta]\}^{-1} [\hat{\underset{\sim}{X}}_* \underset{\sim}{X}_\Delta]' \underset{\sim}{y}$$

Een voorbeeld kan worden ontleend aan model (3.8). Eerder bleek dat dit model overgeïdentificeerd is. De schatting van γ_{11} en γ_{12} is geen probleem indien wordt uitgegaan van de normering $\beta_{21} = 1$. Onder deze voorwaarde kan in vergelijking (3.8.a) η_1 worden vervangen door y_2 waarna OLS schattingen van γ_{11} en γ_{12} kunnen worden verkregen. Dit geldt niet voor β_{31}. Nog steeds uitgaande van $\beta_{21} = 1$, geeft substitutie van (3.8b) in (3.8c)

$$y_3 = \beta_{31} y_2 + \zeta_3^* \tag{3.14}$$

waar $\zeta_3^* = \zeta_3 - \beta_{31} \zeta_2$. OLS is niet mogelijk wegens samenhang van y_2 en ζ_3^*. Zowel x_1 als x_2 is geschikt als instrumentele variabele, hetgeen een van de wijzen is waarop de overidentifikatie zich in de vergelijking (3.8c) openbaart. Een oplossing is regressie van y_2 op x_1 en x_2 met als OLS-schatting y_2. Dit is de eerste fase van de analyse.

Substitutie van $\hat{y}_2$ in (3.14) geeft

$$y_3 = \beta_{31} \hat{y}_2 + \zeta_3^*$$

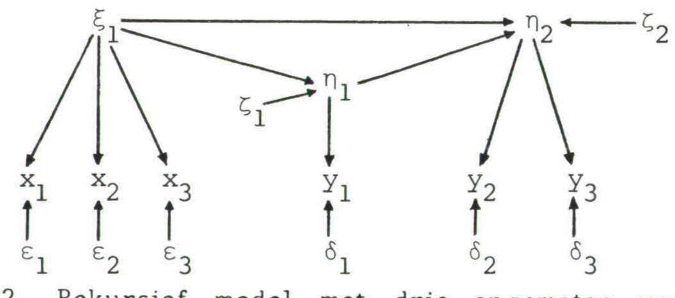
Tijdens de tweede fase van de analyse wordt OLS toegepast op deze vergelijking. Dit geeft

$$\Sigma (y_3\hat{y}_2) = \hat{\beta}_{31} \Sigma \hat{y}_2^2$$

$$\hat{\beta}_{31} = \Sigma (y_3\hat{y}_2)/\Sigma\hat{y}_2^2$$

Voor het terugbrengen van een overtollig aantal instrumentele variabelen is nog een andere oplossing denkbaar dan die via 2SLS. Bedoeld is toepassing van principale komponentenanalyse op basis van alle instrumentele variabelen welke voor vervanging van een bepaalde variabele in aanmerking komen. Als enige instrumentele variabele kiest men dan de komponent met de hoogste eigenwaarde. Nader onderzoek zou moeten uitwijzen welke karakteristieken de schatters hebben die via deze procedure tot stand komen.

Tot slot volgt een schets van een meer uitgebreid rekursief model, geanalyseerd volgens de hierboven uiteengezette "single-equation" procedures.



Figuur 3.2. Rekursief model met drie ongemeten variabelen en zes indikatoren.

De bij dit model behorende vergelijkingen zijn

par.	substitutie	nieuwe vergelijking	instrumentele variabele(n)	IV-schatting	
λ_{x21}	(c) in (d)	$x_2=\lambda_{x21}x_1+\varepsilon^*_2$	x_3	$\hat{\lambda}_{x21}=\Sigma\,(x_2x_3)/\Sigma\,(x_1x_3)$	
λ_{x31}	(c) in (e)	$x_3=\lambda_{x31}x_1+\varepsilon^*_3$	x_2	$\hat{\lambda}_{x31}=\Sigma\,(x_2x_3)/\Sigma\,(x_1x_2)$	
				2SLS-schatting	
				1e fase	2e fase
γ_{11}	(f) en (c) in (a)	$y_1=\gamma_{11}x_1+\zeta^*_1$	x_2, x_3	$\hat{x}_1$: OLS x_1 op x_2, x_3	$\hat{\gamma}_{11}=\Sigma\,(\hat{x}_1y_1)/\Sigma\,(\hat{x}_1^{\,2})$
γ_{21}	(g), (f) en	$y_2=\gamma_{21}x_1$ +	$x_1\rightarrow x_2$, x_3	$\hat{x}_1$: OLS x_1 op x_2, x_3	$\hat{\gamma}_{21}=\dfrac{r(y_1y_2)-r(\hat{x}_1y_2)\,r(\hat{x}_1y_1)}{1-r^2(\hat{x}_1y_1)}\cdot\dfrac{s(y_2)}{s(\hat{x}_1)}$
β_{21}	(c) in (b)	$\beta_{21}y_1+\zeta^*_2$	$y_1\rightarrow x_1, x_2, x_3$	$\hat{y}_1$: OLS y_1 op x_1, x_2, x_3	$\hat{\beta}_{21}=\dfrac{r(\hat{y}_1y_2)-r(x_1y_2)\,r(x_1\hat{y}_1)}{1-r(x_1\hat{y}_1)}\cdot\dfrac{s(y_2)}{s(\hat{y}_1)}$
λ_{y32}	(g) in (h)	$y_3=\lambda_{y32}y_2+\delta^*_3$	x_1, x_2, x_3, y_1	$\hat{y}_2$: OLS y_2 op x_1, x_2, x_3 en y_1	$\hat{\lambda}_{y32}=\Sigma\,(y_3\hat{y}_2)/\Sigma(\hat{y}_2^{\,2})$

hoofdtheorie:

(a) $\eta_1 = \gamma_{11}\xi_1 + \zeta_1$

(b) $\eta_2 = \gamma_{21}\xi_1 + \beta_{21}\eta_1 + \zeta_2$

hulptheorie:

(c) $x_1 = \lambda_{x11}\xi_1 + \varepsilon_1$

(d) $x_2 = \lambda_{x21}\xi_1 + \varepsilon_2$

(e) $x_3 = \lambda_{x31}\xi_1 + \varepsilon_3$

(f) $y_1 = \lambda_{y11}\eta_1 + \delta_1$

(g) $y_2 = \lambda_{y22}\eta_2 + \delta_2$

(h) $y_3 = \lambda_{y32}\eta_2 + \delta_3$

Als gevolg van de rekursiviteit kan in principe de schatting per vergelijking worden afgewerkt. De neiging zal evenwel bestaan een dergelijk model met een moderne FIML-techniek aan te pakken, aangezien hiervoor kant en klare komputerprogramma's voorhanden zijn. Zoals in hoofdstuk 4 is uiteengezet, zijn aan deze techniek naast vele voordelen ook verschillende nadelen verbonden. Het lijkt dan ook nuttig te laten zien hoe genoemd model kan worden geanalyseerd op basis van afzonderlijke vergelijkingen. Uitgaande van de normeringen $\lambda_{x11} = \lambda_{y11} = \lambda_{y22} = 1$ kan de schatting van de zes vrije parameters λ_{x21}, λ_{x31}, λ_{11}, γ_{21}, β_{21} en λ_{y32} als volgt schematisch worden weergegeven. (Zie schema op pag. 83).

Een nadeel van 2SLS is een zekere willekeur in de keuze van normering. Bovendien laat de gelijkstelling van een epistemische parameter aan één zich niet goed in overeenstemming brengen met de interpretatie welke ik in hoofdstuk 5 aan deze parameter geef. Deze interpretatie behelst de aanname dat een epistemische parameter aangeeft de mate waarin een indikator de betekenis van een latente variabele tot uitdrukking brengt. Uit beide opmerkingen volgt dat 2SLS vooral toepasbaar is op modellen met errorvrije- in plaats van latente variabelen. Wil men bovendien de nadelen van "full- information" technieken ontlopen dan wordt een werkwijze zoals in hoofdstuk 6 uitgewerkt, van belang. Hierbij worden de problemen, voortvloeiend uit meetfouten, opgelost via het tot indices verenigen van indikatoren. Vervolgens worden geldigheids- en betrouwbaarheidskoëfficiënten berekend welke als modelparameters in de vorm van restrikties in een struktureel model worden ingevoerd, waarna schatting plaatsvindt van de vrije modelparameters.

3. MODELLEN VAN JORESKOG

In een omvangrijke serie artikelen werd door Jöreskog een aantal algoritmen voor toetsing en schatting van strukturele modellen uitgewerkt, welke speciaal van belang zijn voor de representatie en analyse van theorieën. Er is een drietal hoofdlijnen in het werk van genoemd auteur te onderkennen dat tot deze bewering aanleiding geeft.

Allereerst slaagde de auteur erin algoritmen voor faktor-analytische modellen te ontwikkelen waarmee modellen met ongemeten variabelen en met ingewikkelde niet-rekursieve relatiepatronen kunnen worden onderzocht. Dit was tot voor enkele jaren niet mogelijk. In het vorige hoofdstuk bleek dat faktor-analyse weliswaar gehanteerd kan worden als een struktureel model, maar tevens werd duidelijk dat vele beperkingen aan deze werkwijze zijn verbonden. Gezien vanuit modellen met ongemeten variabelen is de belangrijkste beperking dat geen hoofdtheorie kan worden gespecificeerd. Immers, hoewel de matrix met faktorkorrelaties kan worden geschat, geeft deze geen inzicht in de strukturele relaties tussen faktoren (ongemeten variabelen).

Een eerste uitbreiding kwam in de vorm van een algemeen model voor de analyse van kovariantiestrukturen (Jöreskog 1970). Hiermee is onder meer "second order" faktoranalyse mogelijk, een procedure waarbij faktorkorrelaties worden verklaard door nog verder weg liggende faktoren. Ook kunnen modellen met een rekursieve hoofdtheorie worden onderzocht. Een beperking is echter dat in de hoofdtheorie geen storingstermen zijn toegestaan en dat slechts één indikator per ongemeten variabele in het model kan worden opgenomen. In 1973 kwam dezelfde auteur met een model waarin storingstermen wél zijn toegestaan en weer een jaar later verscheen een artikel waarin modellen met meerdere indikatoren per ongemeten variabele kunnen worden onderzocht (Jöreskog 1973, 1974). Een beperking voor al deze modellen is dat zij gebonden zijn aan de eis van rekursiviteit. In 1976 wordt ook hiervoor een oplossing gegeven in het zogenaamde LISREL-model, tot op heden wel het meest flexibele model op het gebied van modellen met ongemeten variabelen (Jöreskog 1976).

Een tweede ontwikkeling, welke van bepalend belang is geweest voor de empirische ondersteuning van theorieën via de analyse van strukturele modellen, is het ontwerpen van "maximum-likelihood" schattingsprocedures. Hierdoor werd het mogelijk modellen te toetsen op hun "goodness of fit", met behulp waarvan een kontrole op de modelspecifikatie kan worden uitgevoerd. Dit betekent dat een theorie zoals deze is gerepresenteerd in een struktureel model, empirisch kan worden getoetst op juistheid. Een tweede voordeel van kontrole op modelspecifikatie is dat systematische meetfouten kunnen worden onderzocht. Wanneer het vermoeden bestaat dat bepaalde methodenfaktoren in het spel zijn, kunnen deze worden opgenomen in de modelspecifikatie, waarna via een toets op "goodness of fit" de juistheid van dit vermoeden kan worden onderzocht. (Zie voor dergelijke toepassingen vooral Burt 1973 en Long 1976 alsook hoofdstuk 4 van deze dissertatie). Een derde voordeel is dat schattingen niet worden beïnvloed door meeteenheden van indikatoren. Dit laatste is om voor de handliggende redenen voor modellen met latente variabelen van groot belang (zie hoofdstuk 4).
Een laatste pluspunt dat hier te noemen valt, is het feit dat ondanks de bij laatstgenoemde modellen vaak optredende overidentifikatie, efficiënte schatting van modelparameters mogelijk is.

Een derde ontwikkelingslijn waarvoor genoemd auteur van bepalend belang is geweest, is de uitwerking van schattingsmethoden voor modellen ten aanzien waarvan de onderzoeker vooraf modelspecifikaties invoert en als restrikties oplegt aan de parameterschatting. Dit wordt mogelijk gemaakt doordat in een LISREL-analyse naast zogenaamde vrije modelparameters ook *gefixeerde* parameters en *gebonden* parameters mogen voorkomen. Een gefixeerde parameter is een parameter welke voor de aanvang van de schatting een bepaalde waarde wordt toegekend. Gebonden parameters zijn parameters waarvan de waarden een vooraf bepaalde verhouding hebben tot elkaar. Een technisch voordeel van dergelijke parameters is dat een adekwate oplossing kan worden gevonden voor het identifikatieprobleem. Een tweede voordeel is dat specifieke relatiepatronen uit een theorie als een verzameling hypothesen kunnen worden gerepresenteerd in strukturele modellen. Het voor deze studie meest belangrijke voordeel is dat de in hoofdstuk 5 te behandelen betrouwbaarheids- en geldigheidskoëfficiënten als gefixeerde parameters in de analyse van een struktureel model kunnen worden ingevoerd. Die vorm van faktor-analyse waarbij dit laatste mogelijk is, wordt door Jöreskog "*confirmatory*" faktor-analyse genoemd, als onderscheiden van de traditionele "*explorative*" faktor-analyse. In deze "confirmatory" faktoranalyse is bovendien het principe van "maximum-likelihood"-schatting gerealiseerd, waarmee een zeer bruikbare procedure is ontstaan. Om die reden wordt dit type faktormodel hieronder nader toegelicht. Daarna volgt een schets van het tot op heden meest flexibele model, te weten het LISREL-model.

"confirmatory factor analysis" (CFA)

Uitgangspunt is het volgende faktormodel

$$\underset{\sim}{y} = \underset{\sim}{\Lambda}\,\underset{\sim}{f} + \underset{\sim}{u}$$

met als populatie variantie-kovariantie matrix

$$\underset{\sim}{\Sigma}_{yy} = \underset{\sim}{\Lambda}\,\underset{\sim}{\Phi}\,\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}$$

waar $\underset{\sim}{\Sigma}_{yy} = E(\underset{\sim}{y}\underset{\sim}{y}')$, $\underset{\sim}{\Phi} = E(\underset{\sim}{f}\underset{\sim}{f}')$ en $\underset{\sim}{\Psi} = E(\underset{\sim}{u}\underset{\sim}{u}')$ de kovariantie-matrices zijn van respektievelijk inputvariabelen, faktoren en uniciteiten. Stel $\underset{\sim}{T}$ is een willekeurige niet idempotente transformatie-matrix, waarmee de volgende bewerkingen worden uitgevoerd

$$\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\,\underset{\sim}{T}^{-1}\ ,\ \underset{\sim}{\Phi}^* = \underset{\sim}{T}\,\underset{\sim}{\Phi}\,\underset{\sim}{T}',\ \underset{\sim}{f}^* = \underset{\sim}{T}\,\underset{\sim}{f}$$

In dat geval geldt $\underset{\sim}{\Lambda}^*\underset{\sim}{f}^* = \underset{\sim}{\Lambda}\,\underset{\sim}{f}$ en ook $\underset{\sim}{y} = \underset{\sim}{\Lambda}^*\underset{\sim}{f}^* + \underset{\sim}{u}$. We zagen dit probleem al eerder in dit hoofdstuk optreden als gevolg van onbepaaldheid in de meeteenheid van een opgemeten variabele.

Bij een CFA kunnen restrikties worden opgelegd zowel aan faktorladingen als aan faktorkorrelaties. Deze restrikties kunnen betreffen de aanname dat bepaalde parameters een gefixeerde waarde hebben, inklusief de waarde nul, óf dat bepaalde parameters aan elkaar gelijk zijn. Deze mogelijkheid garandeert niet alleen een hoge mate van flexibiliteit ten aanzien van modelspecifikaties, maar kan bovendien leiden tot een meer reëele oplossing van het identifikatieprobleem. Indien voldoende model-assumpties worden geformuleerd, kan de strukturele vorm op unieke wijze uit de gereduceerde vorm worden afgeleid. Er zijn vele regels voor het bepalen van de identifikatiestatus van een model, welke te omvangrijk en komplex zijn om hier te behandelen. Verwezen kan worden naar Anderson en Rubin 1956 en Jöreskog 1969. Voor meer pragmatische procedures van identifikatiekontrole kan worden verwezen naar Werts, Jöreskog en

Linn 1973 alsook een publikatie van de vakgroep methoden en technieken van de vrije universiteit te Amsterdam 1976. (Zie literatuurlijst)

Een ander voordeel van CFA zou kunnen zijn dat de semantische betekenis van faktoren, op te vatten als latente variabelen, vóór de analyse kan worden vastgelegd via model-assumpties. Juist déze eigenschap kan ervoor zorgen dat faktor-analyse uit de exploratieve sfeer komt. Jöreskog hierover: "Factors can be defined to have desired properties and if a preliminary interpretation is available, the restrictions can be chosen to make any subsequent rotation unnecessary". (Jöreskog 1969: 184). Hoewel CFA in vergelijking met de meer traditionele vormen van faktor-analyse op dit punt ontegenzeglijk voordelen biedt, zullen in hoofdstuk 4 enkele bezwaren tegen deze uitspraak worden geformuleerd en uitgewerkt.

Ofschoon CFA zoals reeds is gezegd, niet toestaat de analyse van modellen met een hoofdtheorie, is deze vorm van faktor-analyse van doorslaggevende betekenis geweest voor de ontwikkeling van strukturele modellen met ongemeten variabelen. Dit met name vanwege de mogelijkheid om theoretische inzichten in de vorm van model-assumpties in modellen te representeren. Met de ontwikkeling van CFA is een verbinding tot stand gekomen tussen modellen in de psychometrie en de ekonometrie. Goldberger (1971: 97): "As presented here, the econometric analysis of simultaneous equation models ressembles confirmatory, as distinguished from exploratory factor analysis". Het valt te verwachten dat deze verbinding in de toekomst aanleiding zal zijn tot interessante ontwikkelingen.

lisrel

Hieronder volgt een beschrijving van het type modellen dat met behulp van LISREL kan worden onderzocht. Zoals gezegd is LISREL op het gebied van modellen met ongemeten variabelen wel het meest flexibele systeem dat tot op heden is ontwikkeld. Zo kunnen modellen worden behandeld waarvan de hoofdtheorie een ingewikkeld en niet-rekursief relatiepatroon laat zien, zijn zowel "errors in equations" als "errors in variables" toegestaan en is meervoudige operationalisering mogelijk. In onderstaande wordt het identifikatie- en schattingsprobleem niet aangeroerd, aangezien hiervoor uitstekende handleidingen voorhanden zijn. (Zie literatuurlijst).

Het model betreft G latente endogene variabelen, verzameld in een G*1 vektor $\underset{\sim}{\eta}$ en H latente exogene variabelen, verzameld in een H*1 vektor $\underset{\sim}{\xi}$. Deze latente variabelen vormen gezamenlijk de hoofdtheorie, weer te geven in de volgende verzameling strukturele vergelijkingen.

$$(3.15) \qquad \underset{\sim}{B}\,\underset{\sim}{\eta} = \underset{\sim}{\Gamma}\,\underset{\sim}{\xi} + \underset{\sim}{\zeta} \qquad \text{(hoofdtheorie)}$$

Hier is $\underset{\sim}{B}$ een G*G matrix met strukturele parameters terwijl $\underset{\sim}{\zeta}$ een G*1vektor is met storingstermen. De volgende systeem-assumpties gelden

$$E(\underset{\sim}{\eta}) = E(\underset{\sim}{\xi}) = E(\underset{\sim}{\zeta}) = \underset{\sim}{0}\ ,\ E(\underset{\sim}{\xi}\underset{\sim}{\zeta}') = \underset{\sim}{0}$$

De laatste assumptie betekent dat moet worden aangenomen dat de exogene variabelen niet samenhangen met de storingstermen. De latente variabelen $\underset{\sim}{\eta}$ en $\underset{\sim}{\xi}$ worden gemeten via respektievelijk P indikatoren y_i verzameld in een P*1 vektor $\underset{\sim}{y}$, en Q indikatoren x_i verzameld in een Q*1 vektor $\underset{\sim}{x}$. Gemakshalve wordt ook hier uitgegaan van afwijkingskores. De epistemische relaties kunnen als volgt worden gerepresenteerd. (Variabelen in afwijkingskores)

$$(3.16) \qquad \underset{\sim}{y} = \underset{\sim}{\Lambda}_y\,\underset{\sim}{\eta} + \underset{\sim}{\delta}$$

$$(3.17) \qquad \underset{\sim}{x} = \underset{\sim}{\Lambda}_x \underset{\sim}{\xi} + \underset{\sim}{\varepsilon}$$

$$\underset{\sim}{z} = [\underset{\sim}{y}'\underset{\sim}{x}']'$$

De vektoren $\underset{\sim}{\delta}$ en $\underset{\sim}{\varepsilon}$ betreffen de meetfouten in respektievelijk $\underset{\sim}{y}$ en $\underset{\sim}{x}$, terwijl $\underset{\sim}{\Lambda}_y$ en $\underset{\sim}{\Lambda}_x$ de P*G respektievelijk Q*H matrices zijn met epistemische parameters. Ten aanzien van de meetfouten dienen de volgende systeem-assumpties te worden gemaakt.

$$E(\underset{\sim}{\varepsilon}\underset{\sim}{\eta}')=\underset{\sim}{0} \;,\; E(\underset{\sim}{\delta}\,\underset{\sim}{\eta}')=\underset{\sim}{0} \;,\; E(\underset{\sim}{\varepsilon}\underset{\sim}{\xi}')=\underset{\sim}{0} \;,\; E(\underset{\sim}{\delta}\underset{\sim}{\xi}')=\underset{\sim}{0}$$

$$E(\underset{\sim}{\delta}\underset{\sim}{\zeta}')=\underset{\sim}{0} \;,\; E(\underset{\sim}{\varepsilon}\underset{\sim}{\zeta}')=\underset{\sim}{0} \;,\; E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\delta}')=\underset{\sim}{0}$$

Voorts worden de volgende kovariantie-matrices onderscheiden

$$\underset{\sim}{\Sigma}_{zz}=E(\underset{\sim}{z}\underset{\sim}{z}') \;,\; \underset{\sim}{\Phi}=E(\underset{\sim}{\xi}\underset{\sim}{\xi}') \;,\; \underset{\sim}{\Psi}=E(\underset{\sim}{\zeta}\underset{\sim}{\zeta}')$$

$$\underset{\sim}{\theta}_\varepsilon = E(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}') \;,\; \underset{\sim}{\theta}_\delta = E(\underset{\sim}{\delta}\,\underset{\sim}{\delta}')$$

waar de systeem-assumpties gelden dat $\underset{\sim}{\theta}^2$ en $\underset{\sim}{\theta}^2$ diagonale matrices zijn. Dat wil zeggen, de meetfouten worden geacht onderling niet samen te hangen. De kovariantie-matrix van gemeten variabelen $\underset{\sim}{\Sigma}_{zz}$ kan worden onderverdeeld in drie verschillende submatrices, te weten: $\underset{\sim}{\Sigma}_{xx}$, de P*P matrix met kovarianties tussen indikatoren x_i $i = 1,\ldots,P$, $\underset{\sim}{\Sigma}_{yy}$, de Q*Q matrix met kovarianties tussen indikatoren y_j $j=1,\ldots,Q$ en $\underset{\sim}{\Sigma}_{xy}$, de P*Q matrix met kovarianties tussen x_i en y_j.

$$\underset{\sim}{\Sigma}_{zz} = \begin{pmatrix} \underset{\sim}{\Sigma}_{yy} & \underset{\sim}{\Sigma}_{yx} \\ & \\ \underset{\sim}{\Sigma}_{xy} & \underset{\sim}{\Sigma}_{xx} \end{pmatrix}$$

Met behulp van (3.16) en (3.17) kan deze matrix als volgt worden uitgewerkt

$$(3.18) \quad \underset{\sim}{\Sigma}_{zz} = E \left(\begin{array}{c|c} \underset{\sim}{\Lambda}_y \underset{\sim}{\eta}\underset{\sim}{\eta}' \underset{\sim}{\Lambda}_y' + \underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}' & \underset{\sim}{\Lambda}_y \underset{\sim}{\eta}\underset{\sim}{\xi}' \underset{\sim}{\Lambda}_x' + \underset{\sim}{\varepsilon}\underset{\sim}{\delta}' \\ + \underset{\sim}{\Lambda}_y \underset{\sim}{\eta}\underset{\sim}{\varepsilon}' + \underset{\sim}{\varepsilon}\underset{\sim}{\eta}' \underset{\sim}{\Lambda}_y' & + \underset{\sim}{\Lambda}_y \underset{\sim}{\eta}\underset{\sim}{\delta}' + \underset{\sim}{\varepsilon}\underset{\sim}{\xi}' \underset{\sim}{\Lambda}_x' \\ \hline \underset{\sim}{\Lambda}_x \underset{\sim}{\xi}\underset{\sim}{\eta}' \underset{\sim}{\Lambda}_y' + \underset{\sim}{\delta}\underset{\sim}{\varepsilon}' & \underset{\sim}{\Lambda}_x \underset{\sim}{\xi}\underset{\sim}{\xi}' \underset{\sim}{\Lambda}_x' + \underset{\sim}{\delta}\underset{\sim}{\delta}' \\ + \underset{\sim}{\Lambda}_x \underset{\sim}{\xi}\, \underset{\sim}{\varepsilon}' + \underset{\sim}{\delta}\underset{\sim}{\eta}' \underset{\sim}{\Lambda}_y' & + \underset{\sim}{\Lambda}_x \underset{\sim}{\xi}\underset{\sim}{\delta}' + \underset{\sim}{\delta}\underset{\sim}{\xi}' \underset{\sim}{\Lambda}_x' \end{array} \right)$$

Ook deze matrix kan nog verder worden ontwikkeld met name voor wat betreft $E(\underset{\sim}{\eta}\underset{\sim}{\eta}')$ en $E(\underset{\sim}{\eta}\underset{\sim}{\xi}')$. Met behulp van (3.15) kan voor de kovariantie-matrix van $\underset{\sim}{\eta}$ worden geschreven

$$E(\underset{\sim}{\eta}\underset{\sim}{\eta}') = E\,[(\underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\underset{\sim}{\xi} + \underset{\sim}{B}^{-1}\underset{\sim}{\zeta})\ (\underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\underset{\sim}{\xi} + \underset{\sim}{B}^{-1}\underset{\sim}{\zeta})']$$
$$= \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\,E(\underset{\sim}{\xi}\underset{\sim}{\xi}')\,\underset{\sim}{\Gamma}'\underset{\sim}{B}^{-1\prime} + \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\,E(\underset{\sim}{\xi}\underset{\sim}{\zeta}')\underset{\sim}{B}^{-1\prime}$$
$$+ \underset{\sim}{B}^{-1}E(\underset{\sim}{\zeta}\underset{\sim}{\xi}')\,\underset{\sim}{\Gamma}'\underset{\sim}{B}^{-1\prime} + \underset{\sim}{B}^{-1}E(\underset{\sim}{\zeta}\underset{\sim}{\zeta}')\underset{\sim}{B}^{-1\prime}$$

Aangezien $E(\underset{\sim}{\xi}\underset{\sim}{\zeta}') = \underset{\sim}{0}$, $E(\underset{\sim}{\xi}\underset{\sim}{\xi}') = \underset{\sim}{\Phi}$ en $E(\underset{\sim}{\zeta}\underset{\sim}{\zeta}') = \underset{\sim}{\Psi}$ gaat deze vorm over in

$$E(\underset{\sim}{\eta}\underset{\sim}{\eta}') = \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\,\underset{\sim}{\Phi}\,\underset{\sim}{\Gamma}'\underset{\sim}{B}^{-1\prime} + \underset{\sim}{B}^{-1}\underset{\sim}{\Psi}\,\underset{\sim}{B}^{-1\prime}$$

$$(3.19) \qquad E(\underset{\sim}{\eta}\underset{\sim}{\eta}') = \underset{\sim}{B}^{-1}(\underset{\sim}{\Gamma}\,\underset{\sim}{\Phi}\,\underset{\sim}{\Gamma}' + \underset{\sim}{\Psi})\underset{\sim}{B}^{-1\prime}$$

We herkennen hierin een "second order" faktormodel, evenwel met weglating van storingstermen. (Zie Jöreskog 1974). Voor de kovariantiematrix betreffende $\underset{\sim}{\eta}$ en $\underset{\sim}{\xi}$ geldt naar aanleiding van (3.15)

$$E(\underset{\sim}{\eta}\ \underset{\sim}{\xi}') = \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\,E(\underset{\sim}{\xi}\underset{\sim}{\xi}' + \underset{\sim}{\zeta}\underset{\sim}{\xi}')$$

$$(3.20) \qquad E(\underset{\sim}{\eta}\,\underset{\sim}{\xi}') = \underset{\sim}{B}^{-1}\,\underset{\sim}{\Gamma}\,\underset{\sim}{\Phi}$$

Substitutie van (3.19) en (3.20) in (3.18) geeft tenslotte

$$\Sigma_{zz} = E \left(\begin{array}{c|c} \Lambda_y B^{-1} (\Gamma \Phi \Gamma' + \Psi) B^{-1\prime} \Lambda_y' + \theta_\varepsilon^2 & \Lambda_y B^{-1} \Gamma \Phi \Lambda_x \\ \hline \Lambda_x \Phi \Gamma' B^{-1\prime} \Lambda_y' & \Lambda_x \Phi \Lambda_x' + \theta_\delta^2 \end{array} \right)$$

Het LISREL-programma van Jöreskog voor het schatten van de modelparameters is erop gericht de steekproefkovariantie-matrix S en de uit het model resulterende theoretische kovariantie-matrix Σ_{zz} zo weinig mogelijk te doen verschillen, waarbij de parameters ten aanzien waarvan model-assumpties gelden, worden behandeld als konstanten. Een en ander geschiedt via het op iteratieve wijze minimaliseren van de volgende funktie.

$$\log L = \tfrac{1}{2} N \left[\log|\Sigma_{zz}| + \operatorname{tr} (S \; \Sigma_{zz}^{-1})\right]$$

Voor verdere bijzonderheden aangaande deze procedure wordt verwezen naar de oorspronkelijke literatuur (Jöreskog 1972, Jöreskog and Van Thillo 1974) alsook een publikatie van de vakgroep methoden en technieken van de Vrije Universiteit te Amsterdam 1976).

4. MODELSPECIFIKATIE: EEN KRITIEK

De modellen in hoofdstuk 3 vertonen kenmerken welke bij nadere analyse problematisch blijken te zijn. Allereerst is er sprake van een verstrengeling van een tweetal aspekten: (a) uitspraken op theoretisch nivo worden vertaald in termen van relaties tussen ongemeten variabelen en (b) ongemeten variabelen worden vertaald in termen van indikatoren. De beide hierop betrekking hebbende subsystemen, te weten hoofd- en hulptheorie, worden *tegelijkertijd* in één en dezelfde analyse onderzocht. Zoals in dit hoofdstuk wordt uitgewerkt, ontstaat hierdoor een verwarring van epistemische en strukturele aspekten alsook een verwaarlozing van de problematiek van operationalisering en meting, met als gevolg dat de betrouwbaarheid en geldigheid van theoriekonstruktie onder druk komen.

Een tweede hieronder te behandelen probleem vloeit voort uit het feit dat analyse plaatsvindt op basis van afzonderlijke indikatoren. Voor latente variabelen evenwel geldt dat zij per definitie meerdere betekenisaspekten omvatten, zodat geen der indikatoren de betekenis van zo'n variabele volledig representeert. Wanneer niettemin parameterschatting geschiedt op basis van afzonderlijke indikatoren dan zijn hiervan konsekwenties te verwachten voor de geldigheid van theoriekonstruktie. Het bijzondere van dit geldigheidsprobleem is dat het niet zoals gewoonlijk is gelokaliseerd in de operationaliseringsfase, maar eerst optreedt tijdens de analyse.

Wat betreft de eerstgenoemde problematiek laten nadere analyses de mogelijkheid zien van een uitbreiding van het begrip meetfout naar een konseptie met aanzienlijk meer relevantie voor een politikologisch en sociologisch gebruik van modellen met ongemeten variabelen dan de traditionele opvatting dienaangaande. (Paragraaf 1).

In het kader van de tweede problematiek kunnen zoals zal blijken twee typen "bias" worden onderscheiden, welke nauwe verwantschap vertonen met meetfouten, aangezien zij storende effekten hebben op de betekenisverlening aan latente variabelen. (Paragraaf 2).

Op deze analyses berust een voor deze studie centrale stellingname dat noch de in dit hoofdstuk uitgewerkte typen meetfouten noch de beide typen van "bias", optredend als gevolg van het analyseren op basis van afzonderlijke indikatoren, liggen op het vlak van de traditionele kriteria voor schatting en dat het juister is te spreken van een, overigens juiste, schatting van *niet-beoogde* parameters.

1. MEETFOUTEN

Met de term meetfout wordt vrijwel steeds bedoeld een "random" verstoring van data als gevolg van een onvolkomenheid van (de hantering van) meetinstrumenten. Deze opvatting, welke men tegenkomt in de psychometrie (klassieke meettheorie) en de ekonometrie ("errors in variables") gaat evenwel voorbij aan het bestaan van verstoringen met een systematisch karakter. Zo zijn er systematische meetfouten welke voortvloeien uit het feit dat data niet onafhankelijk van de onderzoeker tot stand zijn gekomen. Deze, tesamen met "random" meetfouten, liggen op het vlak van betrouwbaarheid en worden hier type a meetfouten genoemd. Kenmerkend is dat zij optreden tijdens de fasen van dataverzameling en -verwerking. Daarnaast zijn er systematische meetfouten, liggend op het vlak van geldigheid. Deze vloeien voort uit mankementen in de operationaliseringsfase en worden hieronder type b meetfouten genoemd. Een nadere uitwerking van genoemde typen meetfouten is gewenst.

type a meetfouten

Zoals gezegd betreft het hier in de eerste plaats "random" meetfouten. Voorbeelden zijn vergissingen van de onderzoeker zoals waarnemings-, registratie-, kodeer- en ponsfouten. Ook onderzoekspersonen kunnen zich vergissen in hun reaktie op vragen doordat zij zijn afgeleid of iets verkeerd begrijpen. Aangenomen wordt dat dit soort verstoringen steeds gaan in een willekeurige richting, dat wil zeggen het meetresultaat ofwel de geobserveerde skore wijkt nu eens positief, dan weer negatief af van de werkelijke skore. Dergelijke random fouten worden voortaan type a_1 meetfouten genoemd.

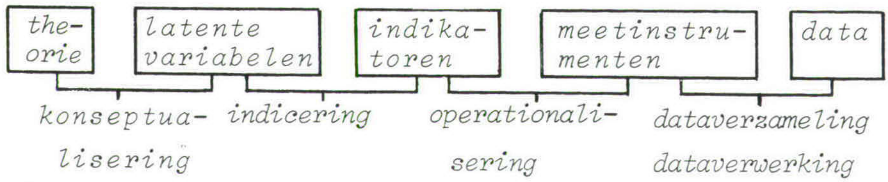
Daarnaast zijn er verstoringen met een systematisch karakter welke evenals type a_1 meetfouten liggen op het vlak van de meetbetrouwbaarheid. Het betreft hier effekten of eigenschappen van indikatoren welke niet zozeer te maken hebben met de inhoud ervan maar met bepaalde bijkomende aspekten. Een bekend voorbeeld is het anticiperen van onderzoekspersonen op heersende meningen en verwachtingspatronen. In dit verband kunnen worden genoemd interviewerbias en sociale wenselijkheid van antwoorden. Een ander voorbeeld vormen leer- en geheugenprocessen. Tijdens het onderzoek leren onderzoekspersonen zich op een bepaalde wijze gedragen of herinneren zich eerdere antwoorden en richten zich op de konsistentie in hun antwoorden.

Ook van de onderzoeker zelf kunnen systematische verstoringen uitgaan tijdens de dataverzameling en -verwerking. Een bekend voorbeeld is het halo-effekt. Hier gaat men er a priori vanuit dat de onderzoekspersonen bepaalde zaken als samenhangend beschouwen. In dit verband spreekt men wel van impliciete persoonlijkheidstheorieën (Barman and Kenny 1976).

De genoemde faktoren met systematische effekten komen overeen met wat Campbell en Fiske *methodenfaktoren* noemen. (Campbell and Fiske 1956). Het voor deze studie meest belangrijke gevolg van een methodenfaktor is dat indikatoren sterker met elkaar korreleren dan bij afwezigheid van zo'n faktor. Het meest duidelijk doet zich dit voor bij het halo-effekt, om welke reden men hier wel spreekt van korrelatiebias. "Nonrandom" meetfouten van dit type worden voortaan aangeduid als type a_2 meetfouten.

type b meetfouten

Vooral wanneer het gaat om latente variabelen is er sprake van een omvangrijk en komplex meetproces. Schematisch kan dit proces als volgt worden weergegeven.



In dit schema zijn een viertal nivo's onderscheiden in een steeds verdergaande konkretisering van een abstrakte latente variabele in de richting van onderzoeksdata. Om van het ene nivo naar het andere te komen is een reeks aktiviteiten nodig, welke aangeduid wordt door de termen indicering, operationalisering, dataverzameling en dataverwerking. Verstoringen welke optreden bij de indicering en operationalisering hebben steeds een systematisch karakter en liggen op het vlak van geldigheid. Ik spreek hier voortaan van type b meetfout. De kern van dit type meetfout is de eis dat tijdens de fase van indicering en operationalisering de betekenis van een theoretisch begrip behouden blijft en tevens dat geen andere betekenissen worden toegevoegd. Deze eis laat zich konkretiseren in het volgende tweetal vragen: (1) wordt een theoretisch begrip in voldoende mate gerepresenteerd door het samenstel van indikatoren en (2) indiceren deze laatste één en slechts één theoretisch begrip? Op grond hiervan ontstaat de volgende tweedeling van type b meetfouten. Indien en voorzover de eerste vraag ontkennend wordt beantwoord, wordt van type b_1 meetfouten gesproken. Eenzelfde antwoord op de tweede vraag geeft aanleiding om te spreken van type b_2 meetfouten. Bovenstaande leidt tot de volgende typologie.

meetfouten	*betrouw-baarheid*	*geldig-heid*
random	a_1	–
nonrandom	a_2	$b \genfrac{}{}{0pt}{}{\nearrow b_1}{\searrow b_2}$

Hieronder wordt de vraag uitgewerkt of, en zo ja in welke richting en mate, modelparameters worden verstoord door de onderscheiden typen meetfouten.

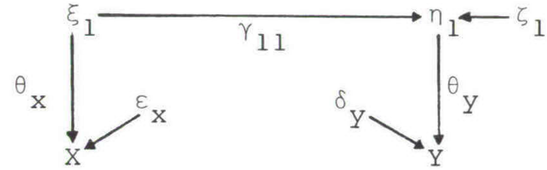
Aangaande random meetfouten kan worden gesteund op een lange traditie in de psychometrie. Het gebied van nonrandom meetfouten evenwel is weinig ontgonnen, dit terwijl juist hier de belangrijkste problemen liggen voor de analyse van modellen met ongemeten variabelen.

verstoring van parameters

De relevantie van de problematiek van random meetfouten (type a_1) in het algemeen en voor kausaal onderzoek in het bijzonder, werd en wordt niet altijd onderkend. Sommigen hanteren het argument dat indien een gekorrigeerde korrelatiekoëfficiënt statistisch signifikant is, dit ook geldt voor een niet-gekorrigeerde koëfficiënt. Dit is onjuist aangezien kan worden bewezen dat een geobserveerde *partiële* korrelatiekoëfficiënt zowel een onder- als een overschatting kan geven van de werkelijke (errorvrije) partiële korrelatiekoëfficiënt (zie Bohrnstedt and Carter 1971). Ook Blalock (1961: 149) liet zien dat zonder korrektie geen zinvolle kausale interpretaties van partiële korrelaties mogelijk zijn.

Een andere misvatting aangaande random meetfouten is dat zij elkaar zouden neutraliseren. Een tweetal overwegingen maakt deze aanname uiterst riskant. Allereerst kan een nivellerend effekt slechts het midden van een meetschaal gelden. Voor extreme waarden zullen verstoringen systematisch in de richting van het schaalmidden gaan, aangezien zij niet of moeilijk nog extremer kunnen worden (zie Bogue and Murphy 1964). Hier is sprake van een *plafond-effekt*. Het gevolg is dat random meetfouten *steeds* een verstoring teweeg brengen in de multivariate verdeling van variabelen (Bogue and Murphy 1964). Dit betekent dat parameters van deze verdeling (varianties, korrelatiekoëfficiënten) systematisch zullen worden verstoord. Het kan worden bewezen dat deze verstoringen gaan in willekeurige richting, hoewel meestal de parameters zullen worden gedrukt. (Bohrnstedt and Carter 1971). Samengevat spelen random meerfouten een dubbele rol: Op univariaat nivo leiden zij tot een verstoring van de verdeling van werkelijke skores en op multivariaat nivo ontstaat (meestal) een reduktie van parameters in de richting van de waarde nul.

Ook type b_1 meetfouten kunnen de analyse van modellen met ongemeten variabelen in sterke mate verstoren. Dit wordt verduidelijkt aan de hand van het model in figuur 4.1. De parameters θ_x en θ_y



Figuur 4.1. Model met 2 ongemeten variabelen, elk gemeten via een index.

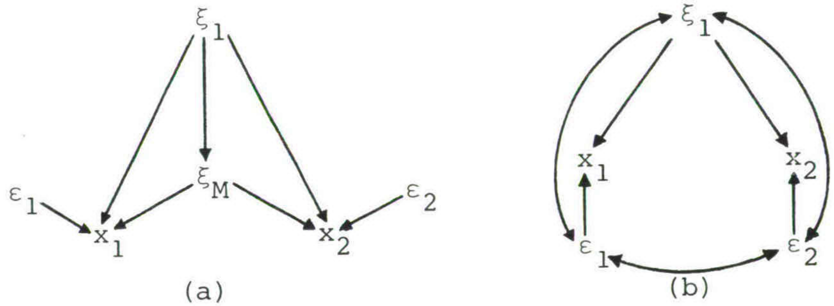
geven een indikatie van de mate van type b_1 meetfouten in respektievelijk X en Y. (Zie voor verdere verklaring paragraaf 1 en 2 van hoofdstuk 5). Stel dat voor dit model geldt $r(XY)=.24$, $\theta_x=.90$, $\theta_y=.89$. Voor γ_{11} wordt dan gevonden $\gamma_{11}= .24/(.90*.89) = .30$. Stel voorts dat type b_1 meetfouten niet zouden zijn onderkend. Zoals in hoofdstuk 5 wordt uiteengezet betekent deze aanname $\theta_x = \theta_y = 1$. In dat geval zou gelden $\gamma_{11}=r(XY)=.24$. Het valt niet moeilijk in te zien dat type b_1 meetfouten steeds zorgen voor een *onderschatting* van *strukturele* parameters.

Type a_2 en b_2 meetfouten kunnen, zowel voor wat betreft hun opsporing (hoofdstuk 5) als wat betreft de verstoringen welke zij teweeg brengen, onder één noemer worden gebracht. In beide gevallen is sprake van het meten van iets terwijl het geen objekt is van meting. Er zijn echter grote verschillen in aard en konsekwenties van verstoringen.

Type a_2 meetfouten kunnen worden verduidelijkt aan de hand van een voorbeeld betreffende de meting van leeftijd

(ξ_1). Stel dat deze errorvrije variabele wordt gemeten via de vraag (indikator) "wat is uw geboortejaar?" zowel gesteld in een enquête (x_1) als in een interview (x_2).

In zo'n situatie is het mogelijk dat de ouderen systematisch een onderschatting geven en de jongeren een dito overschatting. Het is duidelijk dat een korrelatie ontstaat tussen de errorvrije variabele leeftijd ofwel de werkelijke leeftijd en de meetfout. We kunnen dit ook zien als het gevolg van de werking van een methoden faktor ξ_M. Deze is te omschrijven als "sociale wenselijkheid". De ouderdom is vaak niet *meer* in tel en de jeugd *nog* niet. Deze methodenfaktor opereert niet zelfstandig, maar is in zijn werking afhankelijk (over- versus onderschatting) van de werkelijke leeftijd. Bovenstaande kan worden weergegeven middels model (a) in figuur 4.2.



Figuur 4.2. Een tweetal modellen met nonrandom meetfouten. Model (b) ontstaat wanneer de methodenfaktor ξ_M in model (a) niet wordt onderkend.

De bij model (a) behorende vergelijkingen zijn:

(4.1) $$x_k = \lambda_{k1}\xi_1 + \lambda_{kM}\xi_M + \varepsilon_k \qquad k=1,2$$

(4.2) $$\xi_M = \lambda_{M1}\xi_1$$

Het systematische karakter van type a_2 meetfouten blijkt wanneer wordt uitgegaan van een situatie waarin ξ_M niet wordt onderkend zoals het geval is in model (b). In dat geval wordt het effekt van ξ_M opgenomen in de meetfoutkomponenten ε_k k=1,2.

$$(4.3) \quad x_k = \lambda^{\square}_{k1}\xi_1 + \varepsilon^{\square}_k \text{ waar } \varepsilon^{\square}_k = \lambda_{kM}\xi_M + \varepsilon_k \quad k=1,2$$

Hier geeft □ aan dat het gaat om de gewijzigde situatie zonder ξ_M. Het gevolg is dat de basis-assumpties uit de klassieke meettheorie (zie (5.2)) niet langer opgaan. Immers

$$(4.4) \qquad \sigma(\varepsilon^{\square}_k \xi_1) = \lambda_{kM}\lambda_{M1} \neq 0 \qquad k=1,2$$

en

$$(4.5) \qquad \sigma(\varepsilon^{\square}_1 \varepsilon^{\square}_2) = \lambda_{1M}\lambda_{2M} \neq 0$$

Dit leidt ertoe dat model (b) in figuur 4.2 kan worden beschouwd als een gelijkwaardig alternatief voor model (a).

Een tweede gevolg van a_2 meetfouten is een systematische verstoring van de kovariantie tussen twee variabelen x_1 en x_2. Op grond van (4.1) en (4.2.) geldt

$$(4.6) \quad \sigma(x_1x_2) = \lambda_{11}\lambda_{21} + \lambda_{11}\lambda_{2M}\lambda_{M1} + \lambda_{21}\lambda_{1M}\lambda_{M1} + \lambda_{1M}\lambda_{2M}$$

De kovariantie tussen x_1 en x_2 berust kennelijk voor een deel op de werking van een methodenfaktor, welk deel als kunstmatig moet worden bestempeld.

Een andere konsekwentie van het niet onderkennen van ξ_M is dat de parameters λ_{k1} k=1,2 onzuiver worden geschat, als gevolg van het feit dat ξ_1 en $\varepsilon^{\square}_k$ in (4.3.) samenhangen. Ervan uitgaande dat de ongemeten variabelen zijn gegeven in standaardskores en de gemeten variabelen in afwijkingsskores, is $\lambda^{\square}_{k1}$ k=1,2 in (4.3.) gelijk aan de kovariantie tussen x_k en ξ_1.Op basis van (4.1) en (4.2) kan voor deze kovariantie worden geschreven

$$\sigma(x_k \xi_1) = \lambda^{\square}_{k1} = \lambda_{k1} + \lambda_{kM}\, \rho(\xi_1 \xi_M)$$

ofwel

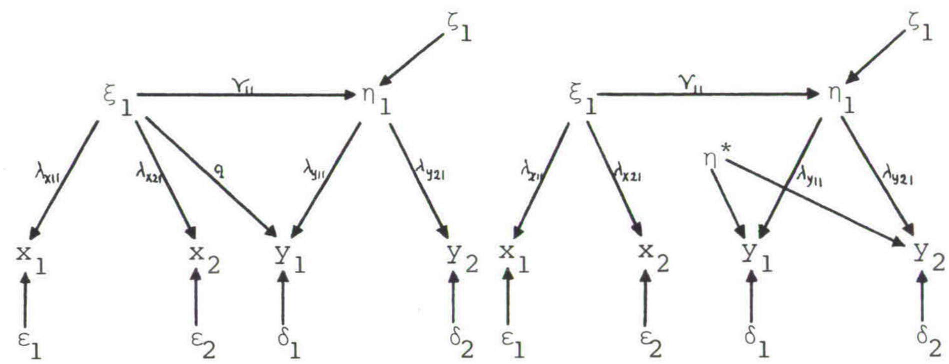
(4.7) $$\lambda^{\square}_{k1} = \lambda_{k1} + \lambda_{kM}\lambda_{M1} \qquad k=1,2$$

Het blijkt dat $\lambda^{\square}_{k1}$ een overschatting geeft van λ_{k1}. Resumerend kan worden gezegd dat de aanwezigheid van een methodenfaktor leidt tot verstoring van zowel de assumpties uit de klassieke meettheorie als van de schatting van epistemische parameters.

Wanneer de methodenfaktor niet samenhangt met de errorvrije variabele, zijn de gevolgen iets minder ingrijpend. In dat geval geldt $\lambda_{M1}=0$ zodat $\sigma(\varepsilon^{\square}_k \xi_1)= 0$ (zie (4.4)). Met andere woorden, een samenhang tussen meetfout $\varepsilon^{\square}_k$ en errorvrije variabele ξ_1 blijft uit. Bovendien volgt uit (4.7) dat de schatting van epistemische parameters niet wordt verstoord.

Alvorens tot slot over te gaan op een behandeling van de konsekwenties van type b_2 meetfouten voor modellen met ongemeten variabelen, wordt gewezen op een in hoofdstuk 5 nader uit te werken wijze voor het opsporen van dergelijke fouten, te weten de analyse van konseptmodellen. De essentie van deze werkwijze is dat indikatoren, bedoeld voor de meting van één (unidimensionele) latente variabele, worden onderzocht op unidimensionaliteit. Brengt de analyse naar voren dat naast de te meten variabele nog andere latente variabelen in het spel zijn, dan zijn er twee mogelijkheden.

Een eerste geval is afgebeeld in model (a) in figuur 4.3., waar een latente variabele ξ_1, welke voor wat betreft de indikator y_1 geen objekt is van meting maar er niettemin een epistemische relatie mee heeft ($q \neq 0$), onderdeel uitmaakt van de hoofdtheorie. Wanneer de epistemische relatie tussen y_1 en ξ_1 niet wordt onderkend (specifikatiefout) heeft dit nadelige konsekwenties voor de schatting van de strukturele parameter γ_{11}. Immers de korrelaties tussen y_1 enerzijds en x_1 en x_2 anderzijds, zullen hoger zijn als gevolg van de betekenissamenhang tussen y_1 en ξ_1. Aangezien het de korrelaties $r(X_1Y_1)$ en $r(X_2Y_1)$ zijn waarop de schatting van γ_{11} mede is gebaseerd, zal een en ander aanleiding geven tot overschatting van γ_{11}. Anders gezegd, de aldus gevonden sterkte van kausale afhankelijkheid tussen ξ_1 en η_1 is voor een deel kunstmatig.



(a) (b)

Figuur 4.3. Een tweetal modellen met type b_2 meetfouten.

Voor een verstoring zoals hier beschreven, gebruikt Costner (1969) de term "differential bias". Als voorbeeld geeft hij de verstoring van de relatie (γ_{11}) tussen sociale klasse (ξ_1) en kriminaliteit (η_1), welke verstoring optreedt als gevolg van het feit dat het aantal arrestaties (y_1) zowel een indikator is van kriminaliteit als van sociale klasse. (Zie model (a) in figuur 4.3., waarbij de betekenis van x_1, x_2 en y_2, zijnde niet-essentieël voor het betoog, in het midden wordt gelaten). Het geschetste probleem staat bekend onder de benaming *"drift of indicators"* alsook *kontaminatie*.

Een tweede denkbaar geval bij de opsporing van type b_2 meetfouten is dat een ongemeten variabele (η^*) welke geen objekt is van meting van bepaalde indikatoren (y_1 en y_2), geen onderdeel uitmaakt van de hoofdtheorie (zie model (b) in figuur 4.3.). Anders gezegd, η^* representeert een ongemeten variabele welke géén theoretische relevantie heeft. Deze zal bijdragen aan de samenhang tussen y_1 en y_2, welk deel van de samenhang ten onrechte aan ξ_1 wordt toegeschreven indien η^* niet wordt onderkend. Hier dreigt overschatting voor λ_{y11} en λ_{y21}. In hoofdstuk 5 wordt bepleit om te korrigeren voor dergelijke invloeden.

2. EPISTEMISCHE EN STRUKTURELE BIAS.

Tijdens de fase van analyse van modellen met ongemeten variabelen ontstaan problemen welke het gevolg zijn van een problematische verhouding tussen ongemeten variabelen en indikatoren. Het centrale probleem is hoe inzicht te krijgen in de relaties tussen de ongemeten variabelen op basis van gekonstateerde verbanden tussen de indikatoren.

epistemische bias

Een van de meest ingrijpende problemen hierbij betreft de schatting van epistemische parameters. Wanneer deze schatting wordt uitgevoerd met behulp van een "full-information"-techniek, ontstaat een vervlechting van epistemische en strukturele aspekten van het model, welke een bedreiging vormt voor de semantische eigenschappen van de variabelen in de hoofdtheorie. Voor een verduidelijking wordt uitgegaan van model (a) in figuur 4.3, met als vooronderstelling dat geen type b_2 meetfouten in dit model aanwezig zijn (q=0). Stel dat de korrelatie-matrix $\underset{\sim}{S}$ met steekproefkorrelaties tussen de indikatoren er voor dit model als volgt uitziet. (De gegevens ontleen ik aan Burt (1973: 152).

$$
\begin{array}{c c}
 & \begin{array}{cccc} x_1 & x_2 & y_1 & y_2 \end{array} \\
\underset{\sim}{S} = \begin{array}{c} x_1 \\ x_2 \\ y_1 \\ y_2 \end{array} & \begin{pmatrix} 1 & .24 & .09 & .21 \\ & 1 & .18 & .22 \\ & & 1 & .37 \\ & & & 1 \end{pmatrix}
\end{array}
$$

Stel voorts dat ξ_1 en η_1 met hun respektievelijke indikatoren worden geanalyseerd als twee afzonderlijke submodellen. (In hoofdstuk 5 spreek ik in zo'n geval van konseptmodellen). Een schatting van de epistemische parameters zou er dan als volgt uitzien.

$$\hat{\lambda}_{x11} = \hat{\lambda}_{x21} = \sqrt{r(x_1 x_2)} = \sqrt{.24} = .49$$

$$\hat{\lambda}_{y11} = \hat{\lambda}_{y21} = \sqrt{r(y_1 y_2)} = \sqrt{.37} = .61$$

Schatting van dezelfde parameters, nu met behulp van een FIML-procedure geeft

$$\hat{\lambda}_{x11} = .45,\ \hat{\lambda}_{x21} = .59,\ \hat{\lambda}_{y11} = .49 \text{ en } \hat{\lambda}_{y21} = .75$$

De verschillen tussen beide schattingen zijn het gevolg van het feit dat de FIML-schatting is gebaseerd op de *gehele* korrelatiematrix terwijl de eerste schatting is gebaseerd op twee submatrices van $\underset{\sim}{S}$. De reden waarom bijvoorbeeld in de FIML-schatting $\hat{\lambda}_{x21}$ hoger is dan $\hat{\lambda}_{x11}$ is dat x_2 sterker samenhangt met respektievelijk y_1 en y_2 dan x_1. Een analoge verklaring geldt voor het FIML-resultaat $\hat{\lambda}_{y21} > \hat{\lambda}_{y11}$.

Dat een en ander konsekwenties heeft voor de geldigheid van operationalisering blijkt als wordt nagegaan hoe in het algemeen de betekenisverlening ten aanzien van latente variabelen tot stand komt. De voor sociologen en politikologen meest voorkomende kontekst waarin het probleem van betekenisverlening zoals hier bedoeld zich voordoet, is die van de faktor-analyse. De interpretatie van faktoren (latente variabelen) wordt hier gebaseerd op de betekenis van indikatoren, waarbij faktorladingen als gewichten kunnen worden gebruikt. Dit betekent in het geval van "full information"-schatting dat de betekenis van een latente variabele mede wordt bepaald door indikatoren welke zijn bedoeld voor meting van een andere latente variabele.[1] (Deze laatste noem ik voortaan *externe* indikatoren als onderscheiden van de op de te meten variabele gerichte *interne* indikatoren.) Het gevolg van bovenstaande is dat de geldigheid van operationalisering achteraf wordt aangetast. Dat wil zeggen, ook nadat de fase van operationalisering is afgesloten, worden tijdens de analyse-fase de semantische eigenschappen van latente variabelen beïnvloed, welke beïnvloeding nadelig is voor een zuivere analyse van de hoofdtheorie. Ik stel voor hier voortaan de term *epistemische bias* te hanteren. Evenals het geval was met meetfouten, heeft deze epistemische bias een terugslag op de geldigheid van de hoofdtheorie. Een storing in epistemische parameters heeft een wijziging van strukturele parameters tot gevolg.

Men zou kunnen tegenwerpen dat het geen nadeel is wanneer indikatoren betekenis verlenen aan andere ongemeten variabelen dan die welke zij indiceren. Dit strookt met de opvatting dat theoretische begrippen behalve een empirische ook een nominale betekenis hebben, welke zij ontlenen aan hun relaties met andere begrippen uit dezelfde theorie. (Burt 1973). Boesjes Hommes (1970: 117) spreekt hier van funktionele betekenis. Ik heb een drietal bezwaren tegen deze redenering. Allereerst kan het gestelde hooguit opgaan voor langdurig getoetste theorieën. Voor theorieën welke nog in een beginstadium verkeren, zoals het geval is met de meeste sociologische en politikologische theorieën, is het gevaarlijk al te veel belang toe te kennen aan een nominale betekenisverlening. Maar ook in het geval van een goed ontwikkelde theorie is het argument ongeldig. Te zeggen dat theoretische begrippen hun betekenis mede ontlenen aan andere begrippen uit de theorie is niet hetzelfde als te zeggen dat semantische en nominale betekenis met elkaar moeten worden verward.(Gemakshalve heb ik in deze diskussie even afgezien van het onderscheid tussen theoretische begippen en ongemeten variabelen).

De geschetste problematiek wordt nog versterkt doordat de mate waarin externe indikatoren de betekenis van latente variabelen beinvloeden sterk kan variëren. Bovendien is deze beïnvloeding meestal ondoorzichtig zodat het moeilijk is haar in de konklusies te verdiskonteren. Enige mate van duidelijkheid is echter verkrijgbaar aan de hand van de volgende twee regels. De eerste regel is dat externe indikatoren welke sterk samenhangen met interne indikatoren een grote invloed hebben. De tweede regel is dat externe indikatoren een sterker effekt hebben op de betekenisverlening naarmate de interne indikatoren onderling zwakker korreleren. Deze beide regels volgen rechtstreeks uit de FIML-procedure, waar parameters worden geschat als een funktie van korrelaties tussen *alle* indikatoren. De tweede regel is bovendien op puur mathematische gronden te verklaren. Een lage waarde van een korrelatie tussen x_k en x_l kan tot stand komen als gevolg van een lage en een hoge epistemische parameter behorend bij de relatie tussen x_k en x_l enerzijds en ξ_1 anderzijds, alsook het gevolg van twee middelmatige epistemische parameters. Met andere woorden, er is veel speelruimte voor de epistemische parameters, waardoor externe indikatoren gemakkelijk een storende invloed kunnen uitoefenen op de schatting. In het geval van een sterke korrelatie tussen interne indikatoren is deze speelruimte er niet. Een sterke korrelatie leidt noodzakelijk tot twee epistemische parameters met beide een hoge waarde.

De konklusie is dat voor een zuivere betekenisverlening vooral van belang is dat indikatoren van een bepaalde latente variabele sterke onderlinge samenhangen vertonen, terwijl zij bovendien niet al te sterk dienen te korreleren met externe indikatoren.

strukturele bias

Ook ten aanzien van strukturele parameters doen zich in de gebruikelijke methodiek van analyse van modellen met ongemeten variabelen problemen voor. Hier vindt schatting plaats op basis van afzonderlijke indikatoren. Het is echter duidelijk dat een indikator een onvolledige representatie geeft van een latente variabele. De vraag is of hier niet een schatting gebaseerd op indices is aangewezen, aangezien deze laatste moeten worden gezien als een betere representatie van latente variabelen dan welke afzonderlijke indikator dan ook. Het antwoord op deze vraag hangt ervan af of men relaties tussen latente variabelen kan opvatten als een aggregaat van relaties tussen beider indikatoren. Of is deze scheiding onmogelijk en dienen indikatoren voor één en dezelfde latente variabele te worden opgevat als één geheel. Dit laatste lijkt meer plausibel. Immers een latente variabele kan gedacht worden te zijn opgebouwd uit een aantal betekenis-aspekten, welke door indikatoren worden gerepresenteerd. Een reden te meer is dat indikatoren vaak op zeer ondoorzichtige wijze met elkaar kunnen samenhangen. Uitgaande van de juistheid van laatstgenoemde vizie stel ik voor de afwijking die ontstaat wanneer wordt geschat op basis van afzonderlijke indikatoren in plaats van indices, aan te duiden als *strukturele bias*.

Intussen zou de indruk kunnen zijn ontstaan dat de geschetste problemen kunnen worden ondervangen door ongemeten variabelen simpel te vervangen door indices, om vervolgens het aldus verkregen model te schatten met een geschikte techniek, als betrof het een model met uitsluitend meetfoutloze variabelen. Hoewel deze werkwijze door de meeste politikologische en sociologische onderzoekers wordt gevolgd, kan zij op grond van de resultaten van bovenstaande analyse niet aanbevolen worden. Weliswaar kan op deze manier het gevaar voor epistemische en strukturele bias grotendeels worden ondervangen, maar tegelijkertijd wordt een nieuw probleem toegevoegd, te weten het negeren van de diskrepantie tussen index en ongemeten variabele.

Nu een overzicht is ontstaan van mogelijkheden waarop bij de analyse van modellen met ongemeten variabelen de geldigheid van deze laatste kan worden aangetast, kan de voor deze studie centrale stellingname worden geformuleerd dat, hoewel type a_1, a_2, b_1 en b_2 meetfouten alsook epistemische- en strukturele bias leiden tot een vertekende schatting van modelparameters, deze niet vallen onder de traditionele kriteria voor zuiverheid, efficiëntie en konsistentie. Juister is het te stellen dat de aanwezigheid van genoemde storingen erop duidt dat het model waarvoor schatting plaatsvindt een ander model is dan hetwelk objekt dient te zijn van onderzoek. Voor wat meetfouten betreft betekent dit dat modellen waarin deze meetfouten zich voordoen *andere* modellen zijn dan die welke we pretenderen te onderzoeken. Aangaande de epistemische- en strukturele bias houdt dit in dat we modellen onderzoeken met losse indikatoren terwijl we zouden moeten onderzoeken modellen waarin deze laatste zijn samengevoegd tot indices. Met andere woorden, genoemde verstoringen leiden niet zozeer tot een foutieve parameterschatting, alswel tot een (overigens zuivere, konsistente en efficiënte) schatting van niet-beoogde modelparameters.

Deze konstatering heeft als konsekwentie dat de oplossing voor de in paragraaf 1 en 2 uitgewerkte problemen niet kan worden gevonden in de standaard ekonometrische en psychometrische literatuur en dat nodig is een op deze specifieke problemen afgestemde aanpak van modellen met ongemeten variabelen. Hierover gaat het resterende deel van deze studie.

Dit hoofdstuk resumerend kan het volgende worden gezegd. Naast de gebruikelijke vormen van meetfouten, liggend op het vlak van de betrouwbaarheid van meting, kunnen meetfouten worden onderscheiden, welke te maken hebben met de geldigheid van meting. Deze laatste zorgen voor een systematische verstoring van de meetresultaten en als gevolg daarvan tevens van de schatting van modelparameters. Een tweede kategorie van verstoringen treedt op als gevolg van het uitvoeren van de analyse op basis van afzonderlijke indikatoren. Deze konstatering pleit voor het samenvoegen van indikatoren tot indices, dit evenwel met dien verstande dat het onderscheid tussen deze laatste en ongemeten variabelen blijft gehandhaafd.[2)]

Tot slot de opmerking dat de geschetste problemen rond analyse van afzonderlijke indikatoren zich weliswaar sterk doen gelden in "full information"-schattingsmethoden, maar dat dit niet inhoudt een ontkenning van het belang van deze methoden. Het voordeel ervan is dat van zoveel mogelijk kennis omtrent de empirie (de variabelen) gebruik wordt gemaakt. Dit voordeel blijkt evenwel in het geval van strukturele modellen met ongemeten variabelen in een nadeel om te slaan.

5. DE ANALYSE VAN KONSEPTMODELLEN

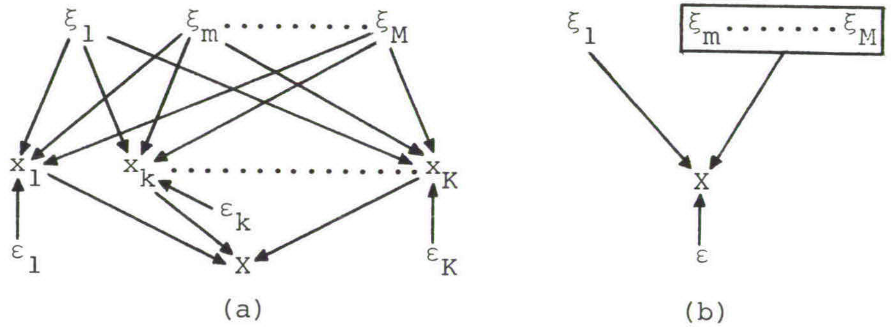
Analyses in hoofdstuk vier leidden tot de konklusie dat verschillende problemen kunnen worden ondervangen wanneer indikatoren worden samengevoegd tot een index. Deze problemen betroffen met name de dreiging van epistemische en strukturele bias. Een ander argument dat pleit voor de konstruktie van indices en dat verder zal worden uitgewerkt in dit hoodstuk, betreft de opsporing van en de korrektie voor de diverse in hoofdstuk 4 onderscheiden typen meetfouten. Met het oog op deze uitwerking wordt in paragraaf 1 een zogenaamd konseptmodel gedefinieerd. Verder wordt een nadere uitwerking gegeven van een tweede konklusie uit hoofdstuk 4, namelijk dat indices en ongemeten variabelen goed van elkaar moeten worden onderscheiden. In paragraaf 2 wordt een procedure uitgewerkt, welke dit onderscheid in kwantitatieve zin mogelijk maakt. Meer in het bijzonder betreft het hier de ontwikkeling van koëfficiënten welke kunnen worden opgevat als epistemische parameters behorend bij de relatie tussen een index en een ongemeten variabele. In paragraaf 3 tenslotte wordt een procedure uitgewerkt voor het konstrueren van indices welke optimaal is in het licht van de in hoofdstuk 4 geanalyseerde meetfoutenproblematiek. Met name gaat het hier om de bepaling van selektiekriteria en wegingsfaktoren voor indikatoren.

1. KONSEPTMODEL

Een voor de analyse van meetfouten geschikt kader is wat voortaan een *konseptmodel* wordt genoemd. Hieronder wordt een verzameling indikatoren x_k $k=1,\ldots,K$ verstaan, welke zijn bedoeld voor de meting van één en slechts één ongemeten variabele ξ_1, alsmede een op basis van deze indikatoren gekonstrueerde index X. Hoewel uitsluitend ξ_1 objekt is van meting, is de kans groot dat de indikatoren naast ξ_1 nog andere ongemeten variabelen indiceren, welke variabelen worden aangeduid als ξ_m $m=2,\ldots,M$. Afhankelijk van hun aard moeten deze ξ_m worden gezien als bronnen van b_2 en of a_2 meetfouten. In het vervolg wordt ten aanzien van *alle* ξ_m $m=1,\ldots,M$ orthogonaliteit verondersteld, dit op grond van de overweging dat samenhangen tussen ongemeten variabelen de hierna uit te werken procedures voor de analyse van meetfouten al te zeer zouden kompliceren.

Konform de sociologische en politikologische onderzoekspraktijk worden indikatoren geïnterpreteerd als effekten van ongemeten variabelen welke zij indiceren. Hieraan ligt impliciet de vooronderstelling ten grondslag dat, indien en voorzover een indikator een expressie dan wel afspiegeling is van een ongemeten variabele, men op hypothetische wijze kan zeggen dat een verandering in deze laatste leidt tot een verandering in de indikatoren. Een en ander steunt op de assumptie van de relatieve belangrijkheid van het betekenisaspekt dat een indikator x_k representeert ten opzichte van de betekenis van de ongemeten variabele ξ_1. Triviaal voor wat betreft de relatie tussen indikatoren en index is de opmerking, dat deze laatste moet worden beschouwd als afhankelijk van de indikatoren.

Op grond van het bovenstaande is het mogelijk een konseptmodel voor te stellen als een model zoals afgebeeld in figuur 5.1.a . Dit model wordt hieronder in een drietal stappen besproken, te weten de analyse van de relatie tussen (1) de indikatoren en de ongemeten variabelen, (2) de indikatoren en de index en (3) de index en de ongemeten variabelen. Het is deze derde stap waarin recht wordt gedaan aan het onderscheid tussen index en ongemeten variabele, op het belang waarvan eerder werd gewezen.



Figuur 5.1. Schematische voorstelling van een konseptmodel.

indikatoren versus ongemeten variabelen

Uit de aard van het geschetste model volgt dat de relaties tussen indikatoren en ongemeten variabelen (1$^{\underline{e}}$ stap) kunnen worden weergegeven via een vergelijking van de volgende vorm

$$x_k=\lambda_{k1}\xi_1 + \sum_{m=2}^{M}\lambda_{km}\xi_m + \varepsilon_k \qquad k=1,\dots,K \tag{5.1}$$

waar K slaat op het aantal indikatoren dat is bedoeld voor de meting van ξ_1 en waar M staat voor het aantal ongemeten variabelen in het konseptmodel. Aan de hand van deze vergelijking kunnen in principe de diverse typen meetfouten nader worden omschreven. Uitgaande van het symbool ε_k ter aanduiding van random meetfouten (type a_1) in de indikatoren x_k, kunnen deze meetfouten volgens de klassieke meettheorie als volgt worden gekenmerkt

$$E(\varepsilon_k)=0 \quad k=1,\ldots,K; \quad E(\varepsilon_k\xi_m)=0 \quad k=1,\ldots,K; \; m=1,\ldots,M$$

(5.2)

$$E(\varepsilon_k\varepsilon_l)=0 \qquad k,l=1,\ldots,K; \; k\neq l$$

Voor wat betreft de overige typen meetfouten lijken de parameters λ_{km} $k=1,\ldots,K$; $m=1,\ldots,M$ van belang. De parameters λ_{k1} $k=1,\ldots,K$ zijn op een of andere wijze gerelateerd aan type b_1 meetfouten. Indien $\lambda_{k1}=0$ betekent dit zonder meer een maximum aan type b_1 meetfouten voor indikator x_k. Indien λ_{k1} de maximale waarde gelijk aan één bereikt, betekent dit afwezigheid van type b_1 meetfouten voor indikator x_k, op voorwaarde dat wordt uitgegaan van standaardskores voor x_k.

Op analoge wijze zijn de parameters λ_{km} $k=1,\ldots,K$; $m=2,\ldots,M$ verbonden aan type b_2 en type a_2 meetfouten. Een λ_{km} groter dan nul slaat op een b_2 dan wel a_2 meetfout, al naar gelang ξ_m betrekking heeft op een ongemeten variabele welke geen objekt is van meting, dan wel op een methodenfaktor.

aggregatie

Hierboven werd gesteld dat $\lambda_{k1}=1$ duidt op de afwezigheid van type b_1 meetfouten in indikator x_k en $\lambda_{km}=0$ $m=2,\ldots,M$ op de afwezigheid van type a_2 dan wel type b_2 meetfouten.

Het valt evenwel niet te verwachten dat λ_{km} $m=1,..,M$ dergelijke extreme waarden zullen aannemen. Voor wat betreft λ_{k1} werd reeds gezegd dat deze aangeeft de mate waarin indikator x_k de ongemeten variabele ξ_1 representeert. Echter x_k representeert in het geval dat ξ_1 een latente variabele is, slechts één van de vele betekenisaspekten, zodat λ_{k1} per definitie kleiner zal zijn dan één.

De konklusie welke hieruit kan worden getrokken is dat b_1 meetfout een begrip is dat niet zozeer van toepassing is op een afzonderlijke indikator, maar veeleer slaat op een *aggregaat* van indikatoren ofwel een index. Als dit juist is, dienen ook type a_2 en b_2 meetfouten te worden gezien op het nivo van een index.

Een en ander impliceert de noodzaak tot aggregatie van parameters λ_{k1} $k=1,..,K$ tot één parameter. Hetzelfde geldt voor λ_{km} $k=1,..,K;$ $m=2,...,M$. De aggregatie dient in dit laatste geval niet alleen over alle K indikatoren worden uitgevoerd, maar tevens over alle M ongemeten variabelen. (Het feit dat de effekten van ξ_m $m=2,...,M$ moeten worden geaggregeerd, is er de oorzaak van dat deze variabelen in model (b) in figuur 5.1 zijn samengevat in een blok). Immers, $\lambda_{km}=0$ geeft weliswaar aan dat ξ_m geen invloed heeft op x_k, maar dit impliceert geenszins dat er geen andere ξ_m's zijn die wél een storende invloed hebben.

In het vervolg zal blijken dat de beide aldus na aggregatie tot stand gekomen parameters kunnen worden gezien als epistemische parameters behorend bij de relatie tussen een index en ξ_1 respektievelijk het aggregaat van ξ_m $m=2,...,M$.

Ook ten aanzien van type a_1 meetfouten is er aanleiding over te stappen van het nivo van indikatoren naar dat van indices, zij het om een andere reden dan het geval was met type a_2 en b_2 meetfouten. Was deze reden hierboven gelegen in het feit dat een latente variabele bestaat uit meerdere betekenisaspekten, welke men tracht te representeren via een verzameling indikatoren, ten aanzien van type a_1 meetfouten is deze reden gelegen in het schatten van de omvang van dit type meetfouten. Deze omvang wordt, zoals bekend is, aangegeven door een betrouwbaarheidskoëfficiënt, waarvan de schatting noodzakelijkerwijs is gebaseerd op het principe van herhaalde meting. Een veel toegepaste vorm hiervan is de test-hertest methode, dat wil zeggen herhaalde meting met hetzelfde meetinstrument. Aangezien in deze studie uitsluitend wordt uitgegaan van materiaal dat is verzameld op één tijdstip, ga ik in het vervolg uit van een ander type herhaalde meting, te weten meting met behulp van een serie indikatoren (zie ook paragraaf 3 van dit hoofdstuk). Op basis hiervan kan een betrouwbaarheidskoëfficiënt worden geschat, welke kan worden gezien als een indikatie van de omvang van type a_1 meetfouten voor een op de genoemde indikatoren gebaseerde index.

Samenvattend blijkt, naast de dreiging van epistemische- en strukturele bias, ook de opsporing en de bepaling van de omvang van de diverse typen meetfouten een reden te vormen voor de konstruktie van indices. Hoewel een meer uitvoerige diskussie dienaangaande wordt uitgesteld tot paragraaf 3 van dit hoofdstuk, volgt hier alvast een eerste aanzet.

indikatoren versus index

Startpunt voor de konstruktie van een index ($2^{\underline{e}}$ stap) is een vergelijking van de vorm

$$(5.3) \qquad X = w_1x_1 + \ldots + w_kx_k + \ldots + w_kx_k$$

De invoering van parameters w_k (een nadere behandeling hiervan volgt in paragraaf 3) *naast* λ_{km} in vergelijking (5.1) is allerminst vanzelfsprekend. Denkbaar is dat voor

w_k in (5.3) wordt genomen $1/\lambda_{k1}$ $k=1,\ldots,K$,maar er zijn redenen om zulks niet te doen. Een eerste argument dat hiertegen pleit is dat de w_k fungeren als wegingsfaktoren in een vergelijking zonder storingsterm. Dit argument loopt parallel aan hetwelk in de literatuur betreffende faktor-analyse wordt aangevoerd tegen de gelijkstelling van faktorskorekoëfficiënten en faktorladingen. De parallel zit in het feit dat λ_{k1} kan worden gezien als een faktorlading, terwijl w_k voldoet aan de omschrijving van een faktorskorekoëfficiënt. Voor een meer uitgebreide argumentatie kan dan ook worden verwezen naar de betreffende literatuur. (Men zie bijvoorbeeld Harman 1970 en Lawley and Maxwell 1971).

Een tweede, in het kader van konseptmodellen nog belangrijkere, reden is dat λ_{k1} onderdeel uitmaakt van een vergelijking waarin naast ξ_1 nog andere (ongewenste) ongemeten variabelen aanwezig zijn. Dit terwijl de vergelijking voor X is afgestemd op de meting van ξ_1. Samenvattend; λ_{k1} zorgt voor een gewichtsverdeling van ξ_m $m=1,\ldots,M$ over x_k $k=1,\ldots,K$, terwijl w_k zorgt voor een gewichtsverdeling van x_k $k=1,\ldots,K$ over uitsluitend X (dus indirekt over ξ_1).

index versus ongemeten variabelen

Met het oog op de relatie tussen X en ξ_m $m=1,\ldots,M$ (3^e stap) tenslotte wordt (5.1) gesubstitueerd in (5.3). Dit geeft

$$(5.4) \qquad X=\sum_{k=1}^{K} w_k\lambda_{k1}\xi_1 + \sum_{k=1}^{K}\sum_{m=2}^{M} w_k\lambda_{km}\xi_m + \sum_{k=1}^{K} w_k\varepsilon_k$$

Deze vergelijking neemt een centrale plaats in bij de hieronder te behandelen procedures voor de bepalingen van de omvang van de diverse typen meetfouten. Er zijn verschillende procedures welke in principe hiervoor kunnen worden benut. Vanuit de klassieke meettheorie zijn bijvoorbeeld verschillende betrouwbaarheidskoëfficiënten ontwikkeld. Deze zijn veelal gebaseerd op het test-hertest-model en voorzover zij berusten op het zogenaamde ekwievalente-vormen-model, dienen forse assumpties ten aanzien van de ekwievalentie van indikatoren worden gesteld. Bovendien laat de klassieke meettheorie niet toe een behandeling van nonrandom meetfouten. Er zijn wél verschillende andere methodieken denkbaar welke bruikbaar zijn voor de opsporing en of kwantificering van type a_2 en b_2 meetfouten. Te noemen zijn Costner's pad-analytische methode, "confirmatory" faktor-analyse (CFA) en de analyse van "multitrait-multimethod" matrices (MTMM). Costners methode is gebaseerd op overgeïdentificeerde pad-modellen. Als gevolg van overidentifikatie kunnen zogenaamde *konsistentiekriteria* worden gesteld, waaruit kan worden opgemaakt of, en zo ja welke typen, nonrandom meetfouten in het spel zijn. Een probleem hierbij is dat geen toetsingskriteria voorhanden zijn. Toepassing van CFA is erop gericht na te gaan of een met type a_2 en b_2 meetfouten gespecificeerd model ten opzichte van de korrelatiematrix van indikatoren voldoende "goodness of fit" laat zien (zie Long 1976), terwijl de MTMM-methode berust op de vergelijking van korrelaties tussen meetresultaten van met meerdere methoden gemeten ongemeten variabelen (zie Campbell and Fiske 1956).

Een nadeel van deze laatste methode is dat wel opsporing, maar geen kwantificering mogelijk is van type a_2 meetfouten, terwijl type b_2 meetfouten in het geheel niet kunnen worden onderzocht. Een nadeel van alle drie genoemde methoden is bovendien dat bij het vooraf specificeren van modellen in sterke mate moet worden vertrouwd op theoretische en voorwetenschappelijke inzichten aangaande de mogelijke rol welke met name type a_2 en b_2 meetfouten spelen.

Als alternatief wordt hier voorgesteld de toepassing van exploratieve faktor-analyse op basis van konseptmodellen.

De hierboven genoemde problemen ten aanzien van type a_1, a_2 en b_2 meetfouten kunnen hiermee goeddeels worden ondervangen. Bovendien ontstaat de mogelijkheid tot kwantificering van type b_1 meetfout. Een en ander wordt in de thans volgende paragraaf uitgewerkt,nadat eerst een op de klassieke meettheorie steunende omschrijving en uitwerking is gegeven van het begrip "theoretische betrouwbaarheid voor indices".

2. SCHATTING OMVANG MEETFOUTEN

Aangaande een kwantificering van type a_1 meetfouten kan zoals reeds eerder is gezegd in belangrijke mate worden gesteund op de klassieke meettheorie. Op grond van deze theorie kan een geschikt uitgangspunt worden geformuleerd in de vorm van een theoretische betrouwbaarheidskoëfficiënt voor indices. Een behandeling van de overige typen meetfouten laat deze theorie evenwel niet toe. Vandaar dat hieronder wordt overgegaan tot interpretatie van een konseptmodel als faktormodel. Aldus kan naast een schatting van de omvang van de overige typen meetfouten tevens een betrouwbaarheidskoëfficiënt worden gekonstrueerd welke superieur is aan die welke in de klassieke meettheorie werden ontwikkeld.

theoretische betrouwbaarheid

Een betrouwbaarheidskoëfficiënt is in de klassieke meettheorie gedefinieerd als de gekwadrateerde korrelatiekoëfficiënt tussen een errorvrije variabele en zijn meetresultaat ofwel de proportie errorvrije variantie in een geobserveerde variabele. Een andere formulering hiervoor, welke vooruitloopt op een verderop uit te werken betrouwbaarheidskoëfficiënt, is de proportie variantie in een geobserveerde variabele (index) welke wordt verklaard door de errorvrije variabele die objekt is van meting. De vraag is wat deze omschrijving betekent op het nivo van een index.

Uitgangspunt is een verzameling indikatoren x_k voor elke waarvan de volgende basisvergelijking uit de klassieke meettheorie kan worden geschreven

$$(5.5) \qquad x_k = \xi_k + \varepsilon_k \qquad k = 1,\ldots\ldots,K$$

Hier staat ξ_k voor de errorvrije komponent van indikator x_k terwijl ε_k slaat op een random meetfout, welke laatste is onderworpen aan de assumpties (5.2). Deze vergelijking geeft de splitsing aan van x_k in een systematische- (ξ_k) en een niet-systematische komponent (ε_k). Konform (5.3) wordt een gewogen additieve index geformeerd

$$X = \xi + \varepsilon$$

waar

$$X = \sum_{k=1}^{K} w_k x_k \,,\quad \xi = \sum_{k=1}^{K} w_k \xi_k \,,\quad \varepsilon = \sum_{k=1}^{K} w_k \varepsilon_k$$

Voor X en ξ kunnen thans de volgende varianties worden geschreven

$$\sigma^2(X) = \sigma^2\{\sum_{k=1}^{K} w_k x_k\} = \sum_{k=1}^{K} w_k^2 \sigma^2(x_k) + \sum\sum_{k \neq 1}^{K} w_k w_1\, \sigma(x_k x_1)$$

$$\sigma^2(\xi) = \sigma^2\{\sum_{k=1}^{K} w_k \xi_k\} = \sum_{k=1}^{K} w_k^2 \sigma^2(\xi_k) + \sum\sum_{k \neq 1}^{K} w_k w_1\, \sigma(\xi_k \xi_1)$$

Op grond van de gelijkheid $\sigma(\xi_k \xi_1) = \sigma(x_k x_1)$ $k,l=1,\ldots\ldots,K;$ $k \neq l$ [welke volgt uit de assumpties in (5.2)] en de gegeven definitie van betrouwbaarheid geldt

$$(5.6) \qquad \rho^2(\xi X) = \frac{\sum_{k=1}^{K} w_k^2 \sigma^2(\xi_k) + \sum\sum_{k \neq 1}^{K} w_k w_1\, \sigma(x_k x_1)}{\sum_{k=1}^{K} w_k^2 \sigma^2(x_k) + \sum\sum_{k \neq 1}^{K} w_k w_1\, \sigma(x_k x_1)}$$

Deze formule is wegens het voorkomen van $\sigma^2(\xi_k)$ niet operationeel, om welke reden ik haar aanduid als *theoretische* betrouwbaarheidskoëfficiënt voor index X.

Het voordeel van de tot een index verenigde indikatoren is dat een schatting kan worden gevonden van $\sigma^2(\xi_k)$. Uitgaande van de assumptie dat de indikatoren ekwievalent zijn (waarover meer in paragraaf 3), ligt het voor de hand de gemiddelde inter-indikator kovariantie te nemen als de meest plausibele benadering van $\sigma^2(\xi_k)$. Immers wanneer x_k en x_l twee (nagenoeg) gelijke skore-vektoren opleveren, zal hun kovariantie (nagenoeg) gelijk zijn aan hun beider variantie, zodat vervolgens de gemiddelde variantie van x_k mag worden verwacht een benadering te zijn van de variantie van de ongemeten variabele ξ_k. De gemiddelde inter-indikator kovariantie is gelijk aan

$$\sum\sum_{k \neq l}^{K} \sigma(x_k x_l)/K(K-1)$$

Na substitutie in (5.6) en gemakshalve uitgaand van een ongewogen index volgt

$$\rho^2(X\xi) = \frac{\sum_{k}^{K} \sum\sum_{k \neq l}^{K} \sigma(x_k x_l)/K(K-1) + \sum\sum_{k \neq l}^{K} \sigma(x_k x_l)}{\sum_{k=l}^{K} \sigma^2(x_k) + \sum\sum_{k \neq l}^{K} \sigma(x_k x_l)}$$

ofwel

(5.7) $$\rho^2(X\xi) = \frac{K/(K-1) \sum\sum_{k \neq l}^{K} \sigma(x_k x_l)}{\sum_{k=l}^{K} \sigma^2(x_k) + \sum\sum_{k \neq l}^{K} \sigma(x_k x_l)}$$

Deze koëfficiënt is gelijk aan de tot op heden verreweg meest gehanteerde betrouwbaarheidskoëfficiënt, te weten koëfficiënt α. Een bekend nadeel van deze koëfficiënt is dat de indikatoren moeten voldoen aan de eis van essentiële τ-equivalentie. Twee indikatoren zijn essentiëel τ-equivalent indien hun errorvrije komponenten (ξ_k)(engels "true scores") slechts met een konstante verschillen. Door Lord en Novick (1974) werd aangetoond dat wanneer hieraan niet is voldaan

α een onderschatting geeft van de werkelijke (theoretische) betrouwbaarheid.

faktormodel als meettheorie

Een aanpak waarin dit bezwaar kan worden ondervangen en waarmee bovendien "nonrandom" meetfouten kunnen worden onderzocht is faktor-analyse.[1] De vraag welke hier eerst moet worden onderzocht is of een konseptmodel kan worden geïnterpreteerd als een faktormodel. Uitgangspunt is

$$\text{(5.8)} \qquad \underset{\sim}{x} = \underset{\sim}{\Lambda}\, \underset{\sim}{\xi} + \underset{\sim}{u}$$

waar $\underset{\sim}{x}$ is een K*1 vektor met indikatoren (inputvariabelen), $\underset{\sim}{\Lambda}$ is een K*M matrix met epistemische parameters (faktorladingen), $\underset{\sim}{\xi}$ is een M*1 vektor met ongemeten variabelen (faktoren) en $\underset{\sim}{u}$ is een K * 1 vektor met uniciteiten. Voorts wordt aangenomen dat $\underset{\sim}{\xi}$ is gestandaardiseerd, dat $\underset{\sim}{x}$ is gemeten in afwijkingsskores en dat de ongemeten variabelen alle orthogonaal zijn.

Ten aanzien van de hierboven gestelde vraag kan thans worden gezegd, dat de assumpties welke gelden voor $\underset{\sim}{u}$ in een faktormodel exakt gelijk zijn aan die voor ε_k bij de analyse van een konseptmodel. Deze zijn voor (5.8)

$$E(\underset{\sim}{u})=\underset{\sim}{0}, \quad E(\underset{\sim}{u}\,\underset{\sim}{u}')= \underset{\sim}{\Psi} \text{ diagonaal}, \quad E(\underset{\sim}{u}\,\underset{\sim}{\xi}')=\underset{\sim}{0}$$

Exakt dezelfde assumpties gelden voor vergelijking (5.1) (zie (5.2)). Voor wat betreft de problematiek van het vinden van wegingsfaktoren w_k werd reeds gezegd dat deze identiek is aan die van het schatten van faktorskorekoëfficiënten in een faktoranalyse.

Een probleem bij een faktor-analytische behandeling van konseptmodellen is nog dat in $\underset{\sim}{u}$ een komponent aanwezig kan zijn welke niet als "random" meetfout kan worden geïnterpreteerd. Deze wordt gevormd door specificiteiten *s*. "Random" meetfouten ε_k in (5.1) slaan op een niet-systematische komponent in de indikatoren x_k $k=1,\ldots,K$, terwijl specificiteiten in zoverre systematisch zijn als zij zorgen voor samenhang van indikatoren met variabelen buiten het faktormodel. Dit impliceert voor een faktor-analytische interpretatie van een konseptmodel de voorwaarde dat specificiteiten een te verwaarlozen rol spelen. (Denkbaar is dat faktor-analyse wordt uitgevoerd op *alle* indikatoren uit het kausale model, om aldus de specificiteiten in $\underset{\sim}{u}$ zoveel mogelijk terug te dringen. Ik kom hier aanstonds op terug) Met andere woorden, onder invloed van de assumptie $\underset{\sim}{u}=\underset{\sim}{\varepsilon}$ ontstaat uit (5.8) een konseptmodel

$$(5.9) \qquad \underset{\sim}{x} = \underset{\sim}{\Lambda}\,\underset{\sim}{\xi} + \underset{\sim}{\varepsilon}$$

waar $\underset{\sim}{\varepsilon}$ staat voor een $K*1$ vektor met type a_1 meetfouten.

Teneinde tot een kwantitatieve aanduiding van de diverse typen meetfouten te komen worden $\underset{\sim}{\xi}$ en $\underset{\sim}{\Lambda}$ als volgt gepartitioneerd.

$$\underset{\sim}{\xi} = [\xi_1 \; \underset{\sim}{\xi}_*']' \text{ en } \underset{\sim}{\Lambda} = [\underset{\sim}{\lambda}_1 \; \underset{\sim}{\Lambda}_*]$$

waar ξ_1 is een ongemeten variabele welke objekt is van meting, $\underset{\sim}{\xi}_*$ is een $(M-1)*1$ vektor met ongemeten variabelen $\underset{\sim}{\xi}m$ $m=2,\ldots,M$ welke *geen* objekt zijn van meting, $\underset{\sim}{\lambda}_1$ is een $K*1$ vektor met ladingen op ξ_1 en $\underset{\sim}{\Lambda}_*$ is een $K*(M-1)$ matrix met ladingen op $\underset{\sim}{\xi}_*$. Op grond van deze partitionering kan voor (5.9) worden geschreven

$$\underset{\sim}{x} = [\underset{\sim}{\lambda}_1 \; \underset{\sim}{\Lambda}_*] \quad [\xi_1 \; \underset{\sim}{\xi}'_*]' + \underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{\lambda}_1 \xi_1 + \underset{\sim}{\Lambda}_* \underset{\sim}{\xi}_* + \underset{\sim}{\varepsilon}$$

Een overgang naar het nivo van een index is mogelijk via substitutie van $X=\underset{\sim}{w}'\underset{\sim}{x}$

(5.10) $$X = \underset{\sim}{w}' \underset{\sim}{\lambda}_1 \xi_1 + \underset{\sim}{w}' \underset{\sim}{\Lambda}_* \underset{\sim}{\xi}_* + \underset{\sim}{w}' \underset{\sim}{\varepsilon}$$

Tot nu toe is uitsluitend aan de orde geweest de vertaling van een konseptmodel in faktor-analytische termen, uitmondend in een vergelijking (5.10) welke exakt gelijk is aan vergelijking (5.4). Thans komt het erop aan maten te vinden, aangevend de bijdragen van ξ_1 respektievelijk $\underset{\sim}{\xi}_*$ aan de variantie in X. Bedoeld zijn de parameters welke aangeven de omvang van type b_1 respektievelijk type a_2 en of b_2 meetfouten, waarvan reeds eerder sprake was. Deze kunnen worden berekend via kwadratering van (5.10) en het nemen van verwachtingswaarden.

$$E(X^2) = \underset{\sim}{w}' \underset{\sim}{\lambda}_1 E(\xi_1^{\,2}) \underset{\sim}{\lambda}'_1 \underset{\sim}{w} + \underset{\sim}{w}' \underset{\sim}{\Lambda}_* \, E(\underset{\sim}{\xi}_* \underset{\sim}{\xi}'_*) \underset{\sim}{\Lambda}'_* \underset{\sim}{w} + \underset{\sim}{w}' E(\underset{\sim}{\varepsilon} \, \underset{\sim}{\varepsilon}') \underset{\sim}{w}$$

Wegens standaardisering van ξ_1 en $\underset{\sim}{\xi}_*$ gaat dit over in

(5.11) $$\sigma^2(X) = \underset{\sim}{w}' \underset{\sim}{\lambda}_1 \underset{\sim}{\lambda}'_1 \underset{\sim}{w} + \underset{\sim}{w}' \underset{\sim}{\Lambda}_* \underset{\sim}{\Lambda}'_* \underset{\sim}{w} + \underset{\sim}{w}' \underset{\sim}{\Psi} \underset{\sim}{w}$$

Deze opsplitsing van varianties kan worden genormeerd naar de totaal te verklaren variantie. Deze is gelijk aan

(5.12) $$\sigma^2(X) = E(\underset{\sim}{w}' \underset{\sim}{x} \; \underset{\sim}{x}' \underset{\sim}{w}) = \underset{\sim}{w}' \underset{\sim}{\Sigma}_{xx} \underset{\sim}{w}$$

waar $\underset{\sim}{\Sigma}_{xx}$ staat voor de K*K kovariantiematrix van indikatoren. Op basis van (5.11) en (5.12) wordt de volgende verdeling van proporties door respektievelijk ξ_1, $\underset{\sim}{\xi}_*$ en $\underset{\sim}{\varepsilon}$ verklaarde variantie verkregen.

$$(5.13) \qquad 1 = \frac{\underset{\sim}{w}'\underset{\sim}{\lambda}_1\underset{\sim}{\lambda}_1'\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}} + \frac{\underset{\sim}{w}'\underset{\sim}{\Lambda}_*\underset{\sim}{\Lambda}_*'\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}} + \frac{\underset{\sim}{w}'\underset{\sim}{\Psi}\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}$$

Een betrouwbaarheidskoëfficiënt werd eerder gedefinieerd als de proportie errorvrije variantie in een meetresultaat (hier index). De vraag is welk deel in (5.13) als zodanig kan worden aangewezen. Een errorvrije variabele werd eerder omschreven als het systematische deel van een meetresultaat. Overeenkomstig deze omschrijving kan als betrouwbaarheidskoëfficiënt worden genomen de proportie systematische variantie in een meetresultaat. Deze proportie is in (5.13) gelijk aan

$$\frac{\underset{\sim}{w}'\underset{\sim}{\lambda}_1\underset{\sim}{\lambda}_1'\underset{\sim}{w} + \underset{\sim}{w}'\underset{\sim}{\Lambda}_*\underset{\sim}{\Lambda}_*'\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}$$

Gebruikmakend van λ_x^2 ter aanduiding van een betrouwbaarheidskoëfficiënt volgt nu

$$(5.14) \qquad \lambda_x^2 = \frac{\underset{\sim}{w}'\underset{\sim}{\Lambda}\,\underset{\sim}{\Lambda}'\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}$$

Deze koëfficiënt blijkt exakt gelijk te zijn aan de door Heise en Bohrnstedt ontwikkelde koëfficiënt Ω. (Heise and Bohrnstedt 1970 formule 43)[2]. Bovendien blijkt λ_x^2 te kunnen worden afgeleid uit de theoretische betrouwbaarheidskoëfficiënt (5.6) door als schatting van indikatorbetrouwbaarheden te nemen de respektievelijke kommunaliteiten. De juistheid van beide beweringen wordt formeel bewezen in een appendix.

Voor de omvang van type b_1 meetfouten lijkt de eerste ratio in (5.13), zijnde de proportie van de variantie in X welke wordt verklaard door ξ_1, een geschikte aanduiding. Deze ratio aanduidend als θ_x volgt uit deze omschrijving

(5.15) $$\theta_x = \frac{\underset{\sim}{w}'\underset{\sim}{\lambda}_1}{\sqrt{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}}$$

Koëfficiënt θ_x bevat in de teller de aggregatie van parameters λ_{k1}, op het belang waarvan reeds eerder werd gewezen. Ook werd eerder gesteld dat deze koëfficiënt kan worden gezien als de epistemische parameter behorend bij de relatie tussen X en ξ_1. De juistheid van deze uitspraak kan als volgt worden bewezen. Bekend is dat de parameter behorend bij een relatie als tussen X en ξ_1 als onderdeel van een konseptmodel en uitgaande van gestandaardiseerde skores, gelijk is aan een korrelatie-koëfficiënt. Dat θ_x gelijk is aan de korrelatie tussen X en ξ_1 ($\rho(X\xi_1)$) blijkt na doorvermenigvuldiging van (5.10) met ξ_1 en het nemen van verwachtingswaarden

$$\sigma(X\xi_1)= \underset{\sim}{w}'\underset{\sim}{\lambda}_1+\underset{\sim}{w}'\underset{\sim}{\Lambda}_*\,\sigma(\underset{\sim}{\xi}_*\xi_1)+\underset{\sim}{w}'\sigma(\underset{\sim}{\varepsilon}\,\xi_1)$$

In deze formule is $\sigma(\underset{\sim}{\xi}_*\xi_1)$ gelijk aan nul wegens de veronderstelde orthogonaliteit. Voorts geldt

$$\sigma^2(X)= \underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}$$

zodat volgt

$$\rho(X\xi_1) = \frac{\underset{\sim}{w}'\underset{\sim}{\lambda}_1}{\sqrt{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}}$$

welke formule exakt gelijk is aan (5.15).

Aangezien type b_1 meetfout een *negatief* begrip is, zou men hier eerder denken aan de maximale proportie door ξ_1 verklaarde variantie verminderd met de feitelijke, dat wil zeggen $1-\rho^2(X\xi_1)$. Met opzet is echter gekozen voor $\theta_x=\rho(X\,\xi_1)$, aangezien deze koëfficiënt bijzonder geschikt blijkt te zijn voor korrektie-doeleinden. (Zie hoofdstuk 6).

Op grond van een analoge redenering lijkt de tweede term in (5.13) een geschikte indikatie te geven van de omvang van type a_2 dan wel b_2 meetfouten

$$(5.16) \qquad \theta_{xm} = \frac{\sqrt{\underset{\sim}{w}' \underset{\sim}{\Lambda}_* \underset{\sim}{\Lambda}_*' \underset{\sim}{w}}}{\sqrt{\underset{\sim}{w}' \underset{\sim}{\Sigma}_{xx} \underset{\sim}{w}}}$$

De parameters λ_x^2, θ_x en θ_{xm} worden in hoofdstuk 6 aangewend voor het analyseren van modellen met ongemeten variabelen. Zij zijn met name geschikt om als gefixeerde parameters ingevoerd te worden in een LISREL-analyse.

Een laatste opmerking in dit kader betreft de rol welke specificiteiten spelen. Eerder werd gesuggereerd faktor-analyse toe te passen op *alle* indikatoren uit het kausale model of, voor zover dit verschil uitmaakt, zelfs op alle gemeten variabelen uit het gehele onderzoek. Dit met het oog op het terugdringen van specificiteiten. Het verdient waarschijnlijk aanbeveling zulks te doen wanneer het gaat om het schatten van $\underset{\sim}{\Lambda}$ in het kader van de berekening van een betrouwbaarheidskoëfficiënt (zie formule (5.14)). Het is immers essentiëel voor het op de voorgestelde wijze berekenen van betrouwbaarheidskoëfficiënten dat uniciteiten kunnen worden opgevat als uitsluitend bestaande uit random meetfouten. Het is echter duidelijk dat voor het bepalen van θ_x en θ_{xm} een faktor-analyse moet worden uitgevoerd binnen de kaders van een konseptmodel.

3. INDEXKONSTRUKTIE

In het bovenstaande was de aandacht vooral gericht op de relatie tussen indikatoren x_k k=1,...,K en ongemeten variabelen ξ_m m=1,...,M, alsook op de relatie tussen index X en ongemeten variabele ξ_1. Rest nog een nadere bestudering van de relatie tussen indikatoren x_k en index X. Met name gaat het hier om de *aard* van de voor opname in een index te selekteren indikatoren en de te volgen selektie-procedure, alsook de bepaling van indikatorgewichten w_k met in het licht van meetfouten gunstige eigenschappen.

selektie van indikatoren

Allereerst de aard van de te selekteren indikatoren. Uit de formele definities voor λ_x^2, θ_x en θ_{xm} volgt: $\lambda_x^2 = \theta_x^2 + \theta_{xm}^2$ In woorden, de betrouwbaarheidskoëfficiënt is gelijk aan de som van de gekwadrateerde b_1- en b_2-geldigheidskoëfficiënten. Hoewel deze gelijkheid formeel opgaat (afgezien van de bovenvermelde speciale rol van specificiteiten), is dit niet het geval in materiëel opzicht. Zoals bekend gaat het bij de bepaling van de betrouwbaarheid van een errorvrije variabele om het principe van herhaalde meting, in deze studie operationeel gemaakt via meting met behulp van een reeks indikatoren. Het is duidelijk dat deze laatste maximaal aan elkaar ekwievalent dienen te zijn en ook zullen zijn, aangezien zij niet worden geselekteerd voor de respresentatie van verschillende betekenisaspekten, maar puur vanuit het principe van herhaalde meting. De konsekwentie hiervan is dat het vinden van een ongemeten variabele ξ_m m=2,...,M in een serie van dergelijke indikatoren meestal slaat op een methodenfaktor, dat wil zeggen op een bron van type a_2 meetfouten. Dit neemt niet weg, dat er ook faktoren ξ_m m=2,...,M kunnen zijn welke het gevolg zijn van semantische trekjes van bepaalde indikatoren.

De zaak verandert zodra in een konseptmodel de te meten variabele een *latente* variabele is. In dat geval dient men te beschikken over een serie maximaal *verschillende* indikatoren. Zij dienen van elkaar te verschillen aangezien zij de uiteenlopende betekenisaspekten van de latente variabelen dienen te representeren. Zij verschillen bij voorkeur *ma-*

ximaal van elkaar indien en voorzover dit een volledige representatie van de latente variabele mogelijk maakt met een minimum aantal indikatoren. De konklusie uit bovenstaande is dat de indikatoren, gericht op de vaststelling van de mate van betrouwbaarheid, in materiëel opzicht verschillen van die welke dienen voor de bepaling van de mate van b_1- en b_2-geldigheid.[3)]

Vervolgens is de selektie van indikatoren aan de orde. Met een errorvrije variabele als meetobjekt kan worden volstaan met de eliminatie van minder betrouwbare indikatoren. Voor wat betreft type a_1 meetfouten is daarbij een probleem dat voor de individuele indikatoren geen betrouwbaarheidskoëfficiënten voorhanden zijn. Te denken valt hier aan hantering van de kommunaliteit van indikatoren als selektiekriterium. Aangaande type a_2 meetfouten ligt het voor de hand indikatoren uit de index te elimineren met een sterke lading op een of meer van de ongemeten variabelen ξ_m $m=2,...M$ en *tevens* een zwakke lading op ξ_1.

Indien een latente variabele objekt is van meting, is een uitvoeriger procedure gewenst. In feite kunnen indikatoren x_k van ongemeten variabelen worden beschouwd als errorvrije variabelen, zodat deze indikatoren op hun beurt zouden dienen te worden gemeten via een serie (maximaal dezelfde) indikatoren. Aldus kunnen x_k $k=1,...,K$ worden gekorrigeerd voor type a_1 en a_2 meetfouten op de hierboven beschreven wijze. Bovendien kunnen betrouwbaarheidskoëfficiënten worden berekend, waarmee het mogelijk is korrelaties tussen indikatoren, *in dit geval indices* , te korrigeren voor attenuatie (hoofdstuk 6). Op basis van gekorrigeerde korrelaties zou vervolgens een faktor-analyse uitgevoerd kunnen worden overeenkomstig de uitwerking in paragraaf 2. Deze werkwijze wordt in hoofdstuk 6 verder uitgewerkt.

De geschetste procedure is tamelijk omslachtig en zal om die reden meestal niet gevolgd worden. Indien men niet anderszins kan beschikken over betrouwbaarheidskoëfficiënten, zal deze stap dan ook veelal moeten worden overgeslagen. Vervolgens kan op basis van de al dan niet gekorrigeerde korrelaties het faktormodel worden geschat. Een korrektie voor type a_2 en b_2 meetfouten is dan tot op zekere hoogte mogelijk via het elimineren van indikatoren met een lage lading op ξ_1 en *tevens* een hoge lading op een der ongemeten variabelen ξ_m $m=2,...,M$.

Ook bij de konstruktie van index-skores spelen meetfouten een rol, in zoverre moet worden gezocht naar gewichten w_k (faktorskorekoëfficiënten) welke zorgen voor een minimale omvang van meetfouten. De index-skores worden aangeduid als X welke worden geacht een schatting te geven van de skores op ξ_1. (N.B.: Deze formulering impliceert de gelijkstelling van X en $\hat{\xi}_1$). Er zijn in het verleden verschillende kriteria alsook methodieken ontwikkeld voor het schatten van faktor-skores welke alle in meer of mindere mate voldoen aan een of meerdere van deze kriteria. Tot op heden kan echter niet worden beschikt over een procedure welke aan alle te stellen eisen optimaal voldoet, (zie Susmilch and Johnson 1975 alsook Alwin 1973), zodat men gedwongen is tot een keuze.

De voor deze studie meest aangewezen kriteria zijn die van optimale validiteit en univociteit, welke samenvallen met de hierboven uitgewerkte kriteria voor respektievelijk b_1- en b_2-geldigheid. Een methode voor de schatting van faktorskores welke aan beide kriteria optimaal voldoet, staat bekend als de regressiemethode (Susmilch and Johnson 1975: 180). Hierbij wordt gezocht naar minimalisering van de variantie van het verschil tussen geschatte skores X en latente variabele ξ_1, ofwel minimalisering van

$$E(X-\xi_1)^2 = E(\underset{\sim}{x}'\underset{\sim}{w}-\xi_1)^2$$

Met het oog op minimalisering wordt de eerste afgeleide van deze variantie op nul gesteld

$$E\,[2\underset{\sim}{x}(\underset{\sim}{x}'\underset{\sim}{w}-\xi_1)] = 2\,[\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}-\underset{\sim}{\lambda}_1] = 0$$

zodat

$$\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w} = \underset{\sim}{\lambda}_1 ; \underset{\sim}{w} = \underset{\sim}{\Sigma}_{xx}^{-1}\underset{\sim}{\lambda}_1$$

Deze gewichten kunnen worden gebruikt in de formule voor θ_x (5.15). Dit geeft

$$\theta_x = \frac{\underset{\sim}{\lambda}_1' \; \underset{\sim}{\Sigma}_{xx}^{-1} \; \underset{\sim}{\lambda}_1}{\sqrt{\underset{\sim}{\lambda}_1' \; \underset{\sim}{\Sigma}_{xx}^{-1} \; \underset{\sim}{\Sigma}_{xx} \; \underset{\sim}{\Sigma}_{xx}^{-1} \; \underset{\sim}{\lambda}_1}}$$

ofwel

$$\theta_x = \sqrt{\underset{\sim}{\lambda}_1' \underset{\sim}{\Sigma}_{xx}^{-1} \; \underset{\sim}{\lambda}_1}$$

De koëfficiënt θ_x, aldus gedefinieerd, moet worden gezien als de meest geschikte koëfficiënt voor het bepalen van b_1-geldigheid, met $\underset{\sim}{\lambda}_1$ en $\underset{\sim}{\Sigma}_{xx}$ vervangen door schattingen. Deze koëfficiënt vervult in hoofdstuk 6 een centrale rol bij het korrigeren van modelparameters.

Aan de hand van het in dit hoofstuk geformuleerde konseptmodel kan zowel de aanpak van de klassieke meettheorie als die welke inherent is aan de analyse van de modellen met ongemeten variabelen uit hoofdstuk 4, als volgt worden bekritiseerd. In de klassieke meettheorie wordt ervan uitgegaan dat (a) de indikatoren één en slechts één ongemeten variabele meten en (b) zij dat alle in eenzelfde mate doen. Dit komt neer op het overslaan van de hierboven geformuleerde *eerste stap* in de analyse van een konseptmodel, dat wil zeggen het voor elke indikator x_k bepalen van de epistemische parameters λ_{km}. Bij de modellen in hoofdstuk 4 daarentegen is sprake van het overslaan van de *tweede en derde stap*, te weten het konstrueren van indices en het schatten van parameters welke de relaties aangeven tussen de indices en de ongemeten variabelen welke zij indiceren. Beide problemen worden ondervangen in een analyse zoals uitgewerkt in dit hoofdstuk. Ook bij deze analyse treden echter komplikaties op, welke zullen worden behandeld in het thans volgende hoofdstuk.

6. ONGEMETEN VARIABELEN: EEN ALTERNATIEF

In voorafgaande hoofdstukken bleek dat op basis van een gewijzigde opvatting omtrent het begrip meetfout, alsook op grond van een kritische beschouwing van de relatie tussen ongemeten variabelen en indikatoren, een fundamentele kritiek mogelijk is op gangbare procedures voor de analyse van modellen met ongemeten variabelen. Meer in het bijzonder werd duidelijk dat niet kan worden volstaan met een analyse op basis van afzonderlijke indikatoren en dat niet zonder meer een index gelijk kan worden gesteld aan een ongemeten variabele.

Bij nader toezien blijken deze konstateringen konsekwenties te hebben welke principieël liggen op het vlak van *modelspecifikatie*. In feite wordt er gezegd dat modellen met ongemeten variabelen dienen te worden gespecificeerd zónder indikatoren en mét voor deze laatste in de plaats tredende indices. Bovendien houdt het beweerde in dat modellen zodanig dienen te zijn gespecificeerd dat er een onderscheid is tussen een index en de ongemeten variabele welke deze beoogt te meten. Een dergelijke specifikatie heeft een tweetal ingrijpende gevolgen voor de *schatting* van modelparameters. Allereerst wordt de matrix van kovarianties tussen indikatoren als basis van schatting vervangen door de matrix van kovarianties tussen indices.

Een voordeel hiervan dat direkt in het oog springt, is de mogelijkheid deze laatste matrix te korrigeren voor de storende effekten van meetfouten. Gebruikmaking van de in hoofdstuk 5 uitgewerkte procedure voor de analyse van konseptmodellen, schept niet alleen de mogelijkheid tot korrektie voor type a_1 meetfouten. Tevens kunnen de overige typen meetfouten worden opgespoord en hun storende effekten eendweegs worden geëlimineerd.

Het bovenstaande procedure-voorstel leidt direkt tot een tweede ingrijpend gevolg voor de schatting van modelparameters, namelijk de noodzaak deze schatting in een tweetal fasen uit te voeren. In een eerste fase vindt de analyse van konseptmodellen plaats, inhoudend de samenstelling van geschikte indices, de schatting van de omvang van de diverse typen meetfouten alsmede de korrektie van de matrix met kovarianties tussen indices voor de storende effekten van type a_1, a_2 en b_2 meetfouten. Het betreft in deze eerste fase een analyse van de hulptheorie van een struktureel model.

De korrektie voor type b_1 meetfouten vindt, zoals zal blijken, in feite eerst plaats in een tweede fase van de analyse. Tijdens deze fase, welke is gericht op de analyse van de hoofdtheorie, worden epistemische parameters welke het resultaat vormen van de eerste analyse-fase, als gefixeerde parameters in het model ingevoerd. Op deze wijze wordt gebruik gemaakt van een van de belangrijkste voordelen van de met name door Jöreskog ontwikkelde schattingsprocedures, te weten de mogelijkheid gefixeerde parameters in de vorm van restrikties op te leggen aan de schatting van de zogenaamde vrije modelparameters. Het is de bedoeling van dit hoofdstuk de geschetste tweefasige analyse nader uit te werken.

1. DE KORREKTIE VOOR MEETFOUTEN

Aan het slot van hoofdstuk 5 werd gesproken over komplikaties welke zich kunnen voordoen bij de analyse van konseptmodellen. Deze betreffen het volgende drietal hieronder te behandelen problemen: (a) type a_2 meetfouten interfereren met de bepaling van de omvang van type a_1 meetfouten, (b) een en dezelfde methodenfaktor heeft een verstorende invloed op twee of meer verschillende konseptmodellen en (c) type a_1 en a_2 meetfouten verstoren de bepaling van de omvang van type b_1 en b_2 meetfouten. De behandeling van een vierde probleem, betreffende de komplexiteit welke optreedt indien latente variabelen meerdimensioneel zijn van karakter, wordt uitgesteld tot de volgende paragraaf.

interferentie

Type a_2 meetfouten interfereren op nadelige wijze met de bepaling van de omvang van type a_1 meetfouten in die zin dat zij zorgen voor een onzuivere berekening van betrouwbaarheidskoëfficiënten.

Uitgangspunt is een konseptmodel met indikatoren x_k $k=1,\ldots,K$ gericht op de meting van *errorvrije* variabele ξ_1. Het gestelde probleem doet zich voor wanneer explorative faktor-analyse naast ξ_1 nog andere faktoren ξ_m $m=2,\ldots,M$ naar voren brengt, waarvan er (minstens) één, zeg ξ_M, een methodenfaktor blijkt te zijn. Immers de kovarianties van de indikatoren x_k $k=1,\ldots,K$, op basis waarvan een betrouwbaarheidskoëfficiënt wordt geschat, zullen gedeeltelijk kunstmatig zijn als gevolg van de storende werking van ξ_M. Hiervoor dient dan ook te worden gekorrigeerd alvorens tot de schatting van een betrouwbaarheidskoëfficiënt kan worden overgegaan. N.b., er wordt *niet* gekorrigeerd voor dat deel in de kovarianties dat tot stand komt via de werking van de overige, zogenaamde ongeldige, ongemeten variabelen ξ_m $m=2,\ldots,M-1$, aangezien deze laatste bijdragen aan een stuk systematisch variantie in de indikatoren dat niet kan worden gekenschetst als artefakt. *Hierin ligt de zin van het van elkaar onderscheiden van type a_2 en b_2 meetfouten, ondanks hun gelijkheid in formeel opzicht.*

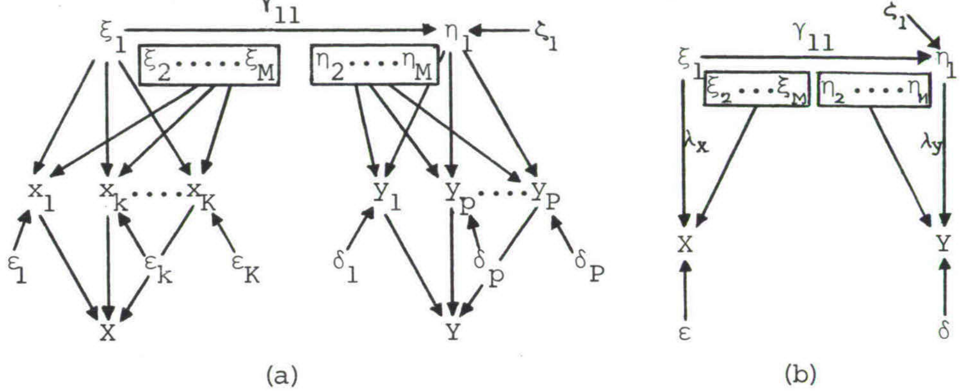
Het hier geschetste geval komt overeen met de situatie zoals weergegeven in de vergelijkingen (4.1) en (4.2) op bladzijde 102. Met gebruikmaking van de vergelijking (4.6) kan de korrektie van de kovarianties tussen indikatoren x_k en x_l $k,l=1,\ldots,K$; $k\neq l$ als volgt geschieden.

$$(6.1)\qquad \sigma^*(x_k x_l) = \lambda_{k1}\lambda_{l1} - (\lambda_{k1}\lambda_{lM}\lambda_{M1} + \lambda_{l1}\lambda_{kM}\lambda_{M1} + \lambda_{kM}\lambda_{lM})$$

waar $\sigma^*(x_k x_l)$ staat voor de kovariantie tussen x_k en x_l welke vrij is van type a_2 meetfouten. Aldus kan een kovariantie-matrix $\underset{\sim}{\Sigma}^*_{xx} = [\sigma^*(x_k x_l)]$ worden verkregen welke geschikt is om te worden gesubstitueerd voor $\underset{\sim}{\Sigma}_{xx}$ in formule (5.14) voor het berekenen van een betrouwbaarheidskoëfficiënt.

koppeling van konseptmodellen

Een andere komplikatie doet zich voor wanneer een en dezelfde methodenfaktor van invloed is op meerdere konseptmodellen. In dat geval is het niet langer mogelijk deze konseptmodellen te behandelen als geheel zelfstandige eenheden. Ter inleiding op deze problematiek wordt als uitgangspunt gekozen een struktureel model van de meest eenvoudige vorm, met slechts twee *errorvrije* variabelen ξ_1 en η_1 in de hoofdtheorie, welke zijn geoperationaliseerd via de respektievelijk indikatoren x_k $k=1,\ldots,K$ en y_p $p=1,\ldots,P$. Een voorbeeld van een dergelijk model is afgebeeld in figuur 6.1. In model (a) staan ε_k $k=1,\ldots,K$ en δ_p $p=1,\ldots,P$ voor



Figuur 6.1. Struktureel model met twee errorvrije variabelen. Model (b) ontstaat uit model (a) via de vervanging van indikatoren door indices.

random meetfouten in x_k en y_p, terwijl deze fouten in model (b) (indexnivo) worden aangegeven door ε en δ. In dit model (b) staan λ_x en λ_y voor de wortels uit de bij X en Y behorende betrouwbaarheidskoëfficiënten. In beide modellen wordt ervan uitgegaan dat exploratieve faktor-analyses op de indikatoren x_k $k=1,\ldots,K$ alsook op de indikatoren y_p $p=1,\ldots,P$, naast ξ_1 en η_1 tevens ongemeten variabelen ξ_m $m=2,\ldots,M$ respektievelijk η_m $m=2,\ldots,M$ naar voren brengen. Ter vereenvoudiging nemen we voorts voorlopig even aan dat géén van deze ongemeten variabelen kan worden aangemerkt als methodenfaktor. De effekten van de genoemde ongeldige variabelen dienen te worden geaggregeerd, hetgeen hun samenvoeging in blokken verklaard. (Zie beide modellen.)

De vraag rijst hoe een model zoals afgebeeld in figuur 6.1.(a) moet worden geanalyseerd. Om te beginnen kunnen ter voorkoming van epistemische- en strukturele bias de indikatoren x_k en y_p worden samengevoegd tot de respektievelijke indices X en Y. Vervolgens kan de korrelatie r(XY) worden berekend als schatting van $\rho(XY)$. Deze laatste kan in het onderhavige geval worden beschouwd als de *gestandaardiseerde* strukturele parameter γ_{11}, waarmee het schattingsprobleem is opgelost.

Hoewel dit de manier is waarop sociologen en politikologen gewend zijn hun schattingsproblemen op te lossen, bleek in hoofdstuk 4 dat hierop een fundamentele kritiek mogelijk is. Daar werd gepleit voor een onderscheid tussen indices en ongemeten variabelen, terwijl in hoofdstuk 5 parameters werden ontwikkeld welke dit onderscheid in kwantitatieve zin mogelijk maken. In het onderhavige geval, waar het gaat om errorvrije variabelen, komen van deze parameters met name in aanmerking de betrouwbaarheidskoëfficiënten λ_x^2 en λ_y^2. De berekening van deze koëfficiënten dient te geschieden binnen twee afzonderlijke konseptmodellen, gecentreerd rond de indikatoren x_k k=1,...,K respektievelijk y_p p=1,...,P.

Door vervolgens de genoemde indikatoren te vervangen door de respektievelijke indices X en Y en door de wortel uit de bijbehorende betrouwbaarheidskoëfficiënten λ_x^2 en λ_y^2 te nemen als epistemische parameters behorend bij de relatie tussen X en ξ_1 enerzijds en tussen Y en η_1 anderzijds, ontstaat model 6.1.(b) uit model 6.1.(a). Voor dit model gelden de volgende modelspecifikaties

hoofdtheorie	hulptheorie
$\eta_1=\gamma_{11}\xi_1+\zeta_1$	(1) $X=\lambda_x\xi_1+\varepsilon$
	(2) $Y=\lambda_y\eta_1+\delta$

Op basis van deze specifikaties kan de strukturele parameter γ_{11} als volgt worden geschat.

$$\rho(XY) = \lambda_x \gamma_{11} \lambda_y$$

zodat

$$\hat{\gamma}_{11} = r(XY)/\hat{\lambda}_x \hat{\lambda}_y \tag{6.2}$$

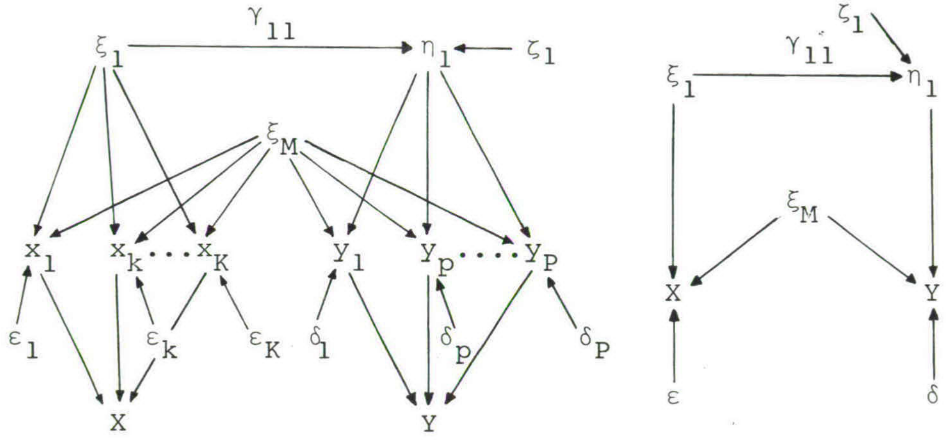
waar $\hat{\lambda}_x$ en $\hat{\lambda}_y$ kunnen worden verkregen met behulp van formule (5.14), nadat hierin parameters zijn vervangen door geschatte waarden.

In het geval ξ_1 en η_1 in de hoofdtheorie niet worden verbonden door de strukturele parameter γ_{11} maar door de simpele nulde-orde korrelatiekoëfficiënt $\rho(\xi_1\eta_1)$, ontstaat een vertrouwd beeld. Immers in dat geval staat $\rho(\xi_1\eta_1)$ voor de korrelatie tussen X en Y welke per definitie vrij is van type a_1 meetfouten. Een voor type a_1 meetfout gekorrigeerde korrelatiekoëfficiënt markerend met een sterretje, kan worden geschreven

$$\hat{\rho}(\xi_1 \eta_1) = r^*(XY) = r(XY)/\hat{\lambda}_x \hat{\lambda}_y \tag{6.3}$$

Bovenstaande staat in de literatuur bekend als korrektie voor attenuatie. N.B. indien ξ_1 en η_1 in bovenstaande duiden op *latente* variabelen, dan heeft dit in formeel opzicht voor bovenstaande uitsluitend de konsekwentie dat de betrouwbaarheidskoëfficiënten λ_x en λ_y worden vervangen door de b_1 geldigheidskoëfficiënten θ_x en θ_y.

In het bovenstaande werd steeds verondersteld dat geen methodenfaktoren in het spel zijn. Thans komt het eigenlijke doel van het voorafgaand betoog aan de orde, namelijk het op bevredigende wijze analyseren van een struktureel model waarin één en dezelfde methodenfaktor een storende invloed heeft op twee of meer konseptmodellen. Met het oog op een verduidelijking wordt eenzelfde soort model behandeld als in figuur 6.1., nadat hieruit ter vereenvoudiging zijn weggelaten de ongemeten variabelen ξ_m m=2,..,M en η_m m=2,..,M, en waaraan is toegevoegd de *methodenfaktor* ξ_M (zie model (a) in figuur 6.2.)



(a) (b)

Figuur 6.2. Struktureel model met een methodenfaktor ξ_M.

Om dezelfde reden en op dezelfde wijze als hierboven wordt het model (a) in figuur 6.2. geherspecificeerd als model (b). Hierbij treedt evenwel als komplikatie op dat als gevolg van ξ_M de konseptmodellen voor ξ_1 en η_1 niet geheel gescheiden kunnen worden geanalyseerd. Dit komt tot uiting in de beide hieronder uitgewerkte mogelijkheden voor het korrigeren voor de verstorende werking van ξ_M: (a) de kovarianties tussen x_k k=1,...,K en y_p p=1,...,P worden gekorrigeerd voor type a_2 meetfouten om vervolgens de berekening van de betrouwbaarheidskoëfficiënten λ_x^2 en λ_y^2 te baseren op de gekorrigeerde kovarianties en (b) de omvang van type a_2 meetfouten wordt geschat via de bepaling van $\hat{\theta}_x$ en $\hat{\theta}_y$, waarna de korrelatie tussen deze indices wordt gezuiverd voor type a_2 meetfouten.

ad a. Men kan hier op dezelfde wijze als voorheen gebruik maken van vergelijking 6.1. voor de korrektie van kovarianties tussen x_k $k=1,\ldots,K$. Eenzelfde soort vergelijking kan worden uitgeschreven voor de indikatoren y_p $p=1,\ldots,P$.

(6.4)

$$\sigma^*(x_k x_l)=\lambda_{k1}\lambda_{l1}-(\lambda_{k1}\lambda_{lM}\lambda_{M1}+\lambda_{l1}\lambda_{kM}\lambda_{M1}+\lambda_{kM}\lambda_{lM}) \quad k,l=1,\ldots,K;k\neq l$$

(6.5)

$$\sigma^*(y_p y_q)=\lambda_{p1}\lambda_{q1}-(\lambda_{p1}\lambda_{qM}\lambda_{M1}+\lambda_{q1}\lambda_{pM}\lambda_{M1}+\lambda_{pM}\lambda_{qM}) \quad p,q=1,\ldots,P;$$
$$p\neq q$$

Hieruit resulteren de $K*K$ respektievelijk $P*P$ kovariantie-matrices $\underset{\sim}{\Sigma}^*_{xx}$ en $\underset{\sim}{\Sigma}^*_{yy}$, welke zijn gezuiverd van type a_2 meetfouten. Het probleem waar het hier om gaat, is het vinden van schattingen voor de parameters λ_{kM} $k=1,\ldots,K$ en λ_{pM} $p=1,\ldots,P$. Het is duidelijk dat deze niet binnen de afzonderlijke konseptmodellen voor ξ_1 en η_1 kunnen worden gevonden. Voorgesteld wordt dit probleem in twee fasen op te lossen. Gestart wordt met een exploratieve faktor-analyse op alle x_k $k=1,\ldots,K$ en y_p $p=1,\ldots,P$ indikatoren, om de aldus gevonden parameters (faktorladingen) te hanteren binnen de afzonderlijke konseptmodellen, dat wil zeggen in de vergelijkingen (6.4) en (6.5). Na substitutie van de op deze wijze gevonden kovariantie-matrices $\underset{\sim}{\Sigma}^*_{xx}$ en $\underset{\sim}{\Sigma}^*_{yy}$ in formule (5.14) verloopt de procedure op identieke wijze als hierboven. (Zie (6.2) en (6.3)).

ad b. Een andere methode, welke in tegenstelling tot de vorige geheel en al is gebaseerd op model (b) hierboven, bestaat uit de korrektie van $\rho(XY)$ voor de invloed van ξ_M. Met andere woorden, de korrektie die wordt voorgesteld vindt plaats op index-nivo in plaats van op het nivo van afzonderlijke indikatoren. Ten aanzien van model (b) gelden de volgende modelspecifikaties

Hoofdtheorie Hulptheorie

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1 \qquad (1)\ X = \lambda_x\xi_1 + \theta_{xM} + \varepsilon$$

$$(2)\ Y = \lambda_y\eta_1 + \theta_{yM} + \delta$$

waar γ_{11} staat voor een strukturele parameter, λ_x en λ_y voor epistemische parameters (wortel uit de betrouwbaarheidskoëfficiënten voor X en Y), θ_{xM} en θ_{yM} voor parameters aangevend de sterkte van de invloed van ξ_M en waar ε en δ staan voor random meetfouten in de respektievelijke indices X en Y. Het systematische karakter van type a_2 meetfouten blijkt wanneer ξ_M over het hoofd wordt gezien. In dat geval kan worden geschreven

$$X = \lambda_x\xi_1 + \varepsilon^{\square} \quad \text{waar} \quad \varepsilon^{\square} = \theta_{xM}\xi_M + \varepsilon$$

$$Y = \lambda_y\eta_1 + \delta^{\square} \quad \text{waar} \quad \delta^{\square} = \theta_{yM}\xi_M + \delta$$

Het gevolg is dat er een samenhang optreedt tussen de meetfout-komponenten.

$$\sigma(\varepsilon^{\square}\delta^{\square}) = \theta_{xM}\theta_{yM} \neq 0$$

Het is duidelijk dat de korrektie van $\rho(XY)$ voor type a_2 meetfouten nu niet kan plaatsvinden zoals hierboven. Voor $\rho(XY)$ immers geldt thans

$$\rho(XY) = \lambda_x\ \rho(\xi_1\eta_1)\ \lambda_y + \sigma(\varepsilon^{\square}\delta^{\square})$$

waaruit als korrektieformule volgt

$$(6.6) \qquad \hat{\rho}(\xi_1\eta_1) = r^*(XY) = \frac{r(XY) - \hat{\theta}_{xM}\hat{\theta}_{yM}}{\hat{\lambda}_x\ \hat{\lambda}_y}$$

Een analoge procedure is toepasbaar voor de schatting van γ_{11}.

Intussen is de vraag blijven liggen hoe θ_{xM}, θ_{yM}, λ_x en λ_y moeten worden geschat. Ook hier is het weer onvermijdelijk te starten met een exploratieve faktor-analyse op *alle* indikatoren x_k $k=1,\ldots,K$ en y_p $p=1,\ldots,P$. De aldus geschatte faktorladingen λ_{kM} $k=1,..,K$ en λ_{pM} $p=1,..,P$ kunnen vervolgens worden toegepast in formule (5.16) voor het bepalen van $\hat{\theta}_{xM}$ en $\hat{\theta}_{yM}$. N.B. De kovariantie-matrices van indikatoren, $\underset{\sim}{\Sigma}_{xx}$ en $\underset{\sim}{\Sigma}_{yy}$, worden in dit geval *niet* gezuiverd voor type a_2 meetfouten, aangezien dit een *dubbele* korrektie zou betekenen. Dientengevolge worden $\hat{\lambda}_x$ en $\hat{\lambda}_y$ op de normale wijze berekend, dat wil zeggen via toepassing van formule (5.14).

Formule (6.6) kan in een vorm worden gebracht welke aansluit bij de traditie in de pad-analyse om ook errortermen te standaardiseren. Voor de varianties van de meetfoutkomponenten ε en δ geldt dan

$$\sigma^2(\varepsilon) = 1-\lambda_x^2, \quad \sigma^2(\delta) = 1-\lambda_y^2$$

zodat (6.6) overgaat in

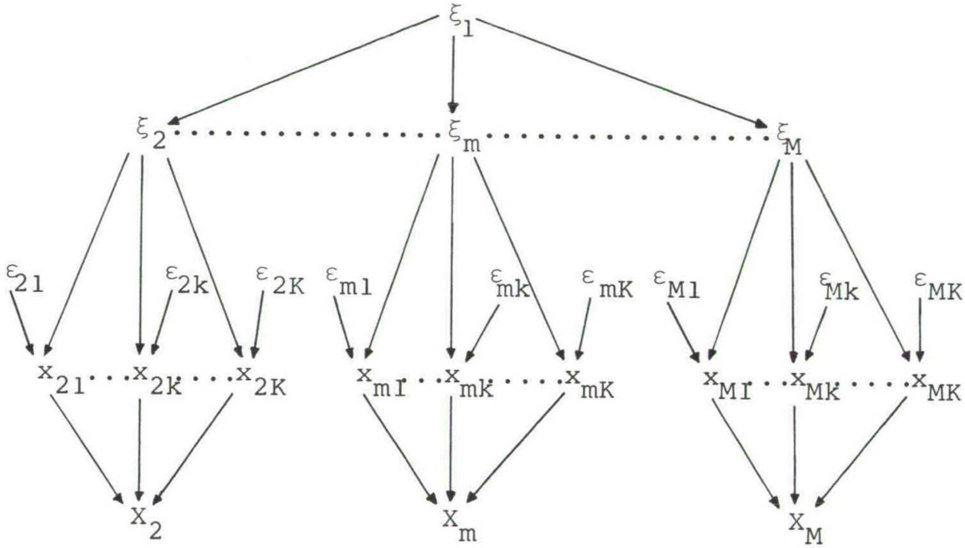
$$r^*(XY) = \frac{r(XY) - \sqrt{(1-\hat{\lambda}_x^2)(1-\hat{\lambda}_y^2)}\,.\; r(\varepsilon^{\square}\delta^{\square})}{\hat{\lambda}_x \quad \hat{\lambda}_y}$$

Deze formule is in formeel opzicht exakt gelijk aan formule 5 van Wiley and Wiley (1973: 175). Een materieel verschil is dat de laatstgenoemde auteurs uitgaan van enkelvoudige variabelen, dit evenwel met als konsekwentie dat zij niet aan kunnen geven hoe $r(\varepsilon^{\square}\delta^{\square})$ alsook λ_x en λ_y moeten worden geschat.

Wanneer een latente variabele objekt is van meting doet zich de eigenaardigheid voor dat voor de meting van een dergelijke variabele een verzameling maximaal verschillende indikatoren nodig is, hetgeen een korrektie voor type a_1 meetfouten in feite onmogelijk maakt. Immers voor de analyse van dit type meetfouten dienen we in principe te beschikken over een verzameling maximaal ekwievalente indikatoren. Dit betekent dat het tegelijkertijd analyseren van alle vier de typen meetfouten op basis van één en dezelfde verzameling indikatoren niet mogelijk is. Het gestelde komt er in feite op neer dat indikatoren voor een latente variabele moeten worden opgevat als errorvrije variabelen, tenzij men op een of andere wijze kan beschikken over betrouwbaarheidskoëfficiënten voor deze indikatoren zonder dat meervoudige operationalisering is uitgevoerd. (Bijvoorbeeld op grond van de resultaten van vergelijkbare onderzoekingen).

Indien men deze forse assumptie betreffende afwezigheid van type a_1 meetfouten in indikatoren van een latente variabele, zeg ξ_1, niet wenst te maken, kan worden gedacht aan de volgende procedure. Stel ξ_1 is een latente variabele, voor de meting waarvan een verzameling maximaal verschillende indikatoren ξ_m $m = 2,\ldots,M$ wordt geselekteerd. Deze indikatoren worden opgevat als errorvrije variabelen, welke op hun beurt een meervoudige operationalisering ondergaan via een verzameling maximaal ekwievalente indikatoren x_{mk} $m = 2,\ldots,M$; $k = 1,\ldots,K$, waar K staat voor het aantal indikatoren dat voor elke errorvrije indikator ξ_m beschikbaar is. (Gemakshalve ga ik uit van een vast aantal (K) indikatoren per ξ_m). Een en ander betekent een forse uitbreiding van de hulptheorie, hetgeen verder kan worden verduidelijkt aan de hand van het model in figuur 6.3.

Aangezien het hier slechts gaat om één latente variabele, kan dit model worden opgevat als een konseptmodel.



Figuur 6.3. Konseptmodel voor een 2-staps analyse van meetfouten.

Voorgesteld wordt om een dergelijk model in de volgende twee stappen te analyseren. In de eerste stap worden op basis van de indikatoren x_{mk} $m=2,\ldots,M$; $k=1,\ldots,K$ betrouwbaarheidskoëfficiënten λ_m^2 $m=2,\ldots,M$ berekend, met behulp waarvan de kovarianties tussen de indikatoren ξ_m $m=2,\ldots,M$ worden gekorrigeerd op de eerder aangegeven wijze. Het resultaat hiervan is de $(M-1)*(M-1)$ kovariantie-matrix $\underset{\sim}{\Sigma}^*_{\xi\xi}$ van de indikatoren ξ_m $m=2,\ldots,M$, welke kan worden gezien als een matrix van kovarianties welke zijn gezuiverd voor de storende effekten van type a_1 meetfouten. Vervolgens kunnen in een tweede stap de b_1-geldigheidskoëfficiënten θ_m $m=2,\ldots,M$ worden berekend via toepassing van formule (5.18), nadat hierin $\underset{\sim}{\Sigma}^{-1}_{xx}$ is vervangen door $\underset{\sim}{\Sigma}^{*-1}_{\xi\xi}$. Het is duidelijk dat indien naast type a_1 en b_1 meetfouten ook type a_2 en b_2 meetfouten een rol spelen, de geschetste werkwijze gemakkelijk kan worden gekombineerd met procedures welke dienaangaande hierboven werden uitgewerkt.

Behalve methodenfaktoren kunnen er nog andere ongemeten variabelen zijn welke ervoor zorgen dat meerdere konseptmodellen tegelijkertijd moeten worden bekeken. Bedoeld zijn variabelen welke geen theoretische relevantie hebben, dus niet zijn opgenomen in de hoofdtheorie, zoals bijvoorbeeld η^* in figuur 4.2. Wanneer een dergelijke variabele een effekt heeft op indikatoren in meerdere konseptmodellen, dan dient voor dit effekt te worden gekorrigeerd op dezelfde wijze als hierboven ten aanzien van type a_2 meetfouten werd gedaan. Het is echter duidelijk dat de effekten van dergelijke theoretisch irrelevante variabelen niet behoeven te worden geëlimineerd alvorens betrouwbaarheidskoëfficiënten worden berekend, zoals het geval is met type a_2 meetfouten. Immers dat deel in een kovariantie dat door toedoen van dergelijke variabelen tot stand komt, kan niet worden gekwalificeerd als een artefakt.

In deze paragraaf werden diverse problemen behandeld welke zich ten aanzien van de analyse van konseptmodellen kunnen voordoen. Deze problemen kunnen een grote diversiteit van verschijningsvormen hebben, hetgeen een behandeling in *algemene* termen schier onmogelijk maakt. Vandaar dat werd gekozen voor een toelichting aan de hand van enkele eenvoudige voorbeelden. Hieruit laat zich gemakkelijk de volgende gedragslijn afleiden. Het is wenselijk om te analyseren binnen zo klein mogelijke eenheden, liefst binnen afzonderlijke konseptmodellen. Zodra er effekten zijn welke dwars door deze kleinere eenheden heenlopen, schat dan de benodigde parameters op basis van de grotere eenheden waarvoor genoemde effekten gelden. Ga vervolgens met de analyse terug naar de kleinere eenheden (konseptmodellen) en maak gebruik van de op basis van grotere eenheden geschatte parameters.

Deze gedragslijn schetst ook de verhouding van de konseptmodellen in deze paragraaf en de strukturele modellen uit de volgende paragraaf.

2. MODELLEN MET RESTRIKTIES.

In onderstaande wordt aan de hand van enkele voorbeelden verduidelijkt hoe strukturele modellen kunnen worden ontleed in konseptmodellen en hoe vervolgens de parameters van deze modellen als restrikties kunnen worden ingevoerd bij de schatting van strukturele parameters. De voorgestelde werkwijze betekent een gescheiden analyseren van hoofd- en hulptheorie.

Allereerst volgt nu de behandeling van een eenvoudig model met drie errorvrije variabelen ξ_1, η_1 en η_2 gemeten via de respektievelijke indikatoren x_{1k} $k=1,\ldots,K$, y_{1p} $p=1,\ldots,P$ en y_{2q} $q=1,\ldots,Q$. Dit model is weergegeven in figuur 6.4.a. Op grond van argumenten die inmiddels voldoende bekend zijn, kan model (b) uit model (a) worden afgeleid door uit de relatie tussen de indices X_1, Y_1 en Y_2 en de errorvrije variabelen ξ_1, η_1 en η_2 de indikatoren weg te laten. Bij de aldus verkregen direkte relaties behoren als epistemische parameters de wortels uit de betrouwbaarheidskoëfficiënten λ^2_{x1}, λ^2_{y1} en λ^2_{y2}. Voor model (b) kan een en ander tot uitdrukking worden gebracht door de volgende specifikaties:

Hoofdtheorie	Hulptheorie
$\eta_1 = \gamma_{11}\xi_1 + \zeta_1$	$X_1 = \lambda_{x1}\xi_1 + \varepsilon_1$
$\eta_2 = \gamma_{21}\xi_2 + \zeta_2$	$Y_2 = \lambda_{y1}\eta_1 + \delta_1$
	$Y_3 = \lambda_{y2}\eta_2 + \delta_2$

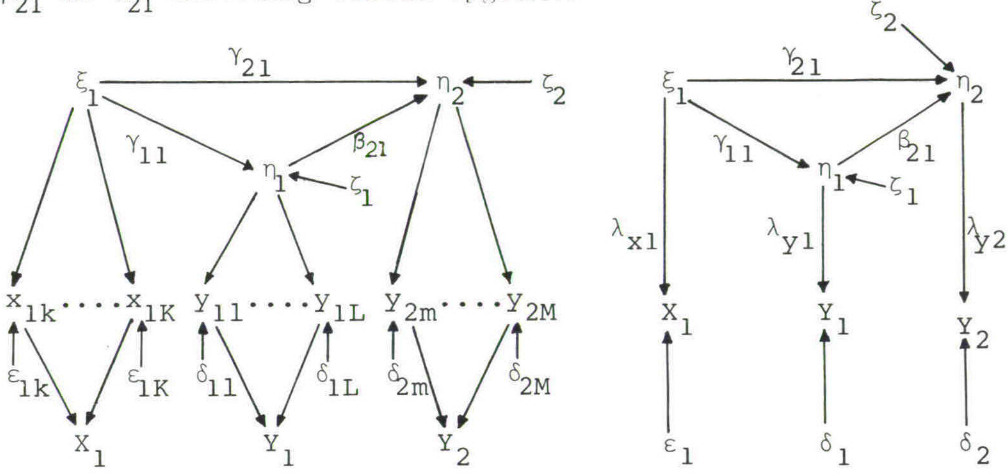
Na substitutie van de vergelijkingen uit de hoofdtheorie in die uit de hulptheorie kunnen de volgende schattingsvergelijkingen worden verkregen

$$\rho(X_1 Y_1) = \lambda_{x1} \gamma_{11} \lambda_{y1}$$

$$\rho(X_1 Y_2) = \lambda_{x1} \gamma_{21} \lambda_{y2} + \lambda_{x1} \gamma_{11} \beta_{21} \lambda_{y2}$$

$$\rho(Y_1 Y_2) = \lambda_{y1} \beta_{21} \lambda_{y2} + \lambda_{y1} \gamma_{11} \gamma_{21} \lambda_{y2}$$

De betrouwbaarheidskoëfficiënten λ^2_{x1}, λ^2_{y1} en λ^2_{y2}, welke op basis van de respektievelijke koseptmodellen voor ξ_1, η_1 en η_2 worden geschat, kunnen als gefixeerde parameters worden uitgevoerd in deze schattingsvergelijkingen. Uit het aldus verkregen stelsel van drie vergelijkingen kunnen vervolgens γ_{11}, γ_{21} en β_{21} eenvoudig worden opgelost.



(a) (b)

Figuur 6.4. Model met errorvrije variabelen ξ_1, η_1 en η_2, gemeten middel de respektievelijke indices X_1, Y_1 en Y_2.

Het is instruktief te laten zien dat een dergelijke procedure in feite kan worden opgevat als gebaseerd op het bekende korrigeren voor attenuatie. Immers, met behulp van de geschatte betrouwbaarheidskoëfficiënten kan worden geschreven

$$r^*(X_1 Y_2) = r(X_1 Y_1)/\hat{\lambda}_{x1} \hat{\lambda}_{y1}$$

$$r^*(X_1 Y_2) = r(X_1 Y_2)/\hat{\lambda}_{x1} \hat{\lambda}_{y2}$$

$$r^*(Y_1Y_2) = r(Y_1Y_2)/\hat{\lambda}_{y1}\hat{\lambda}_{y2}$$

Via substitutie van de gekorrigeerde koëfficiënten r^* in de schattingsvergelijkingen hierboven ontstaat een nieuwe verzameling schattingsvergelijkingen welke we herkennen als behorend bij een model met uitsluitend *gemeten* variabelen.

$$r^*(X_1Y_1) = \gamma_{11}$$

$$r^*(X_1Y_2) = \gamma_{12} + \gamma_{11}\beta_{12}$$

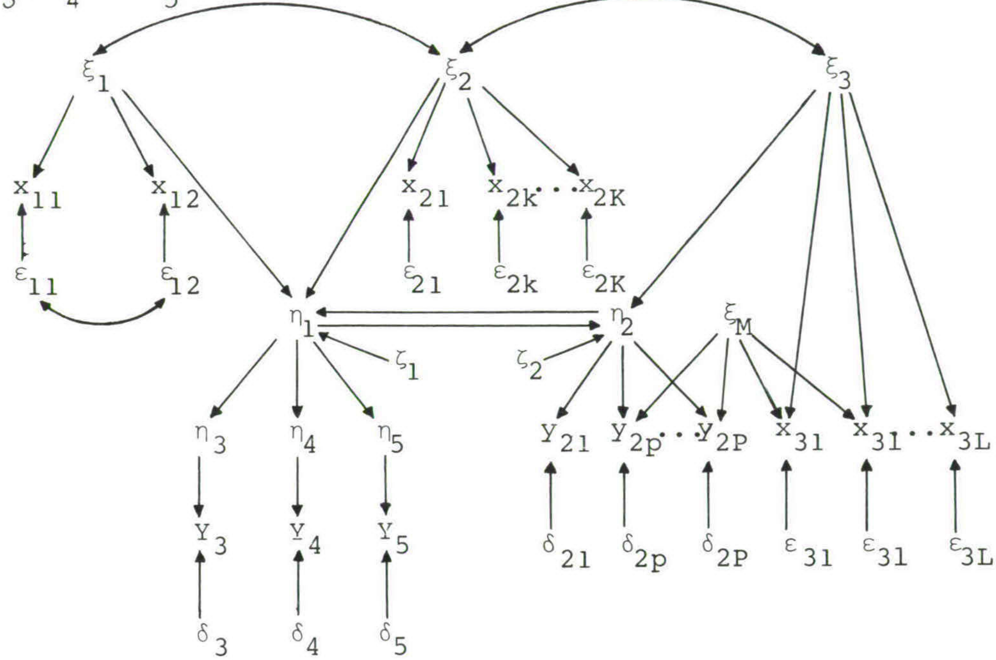
$$r^*(X_1Y_2) = \beta_{12} + \gamma_{11}\gamma_{12}$$

Uit deze vergelijkingen kunnen de onbekenden γ_{11}, γ_{12} en β_{12} eenvoudig worden opgelost.

Beide analyses hierboven tonen duidelijk de onjuistheid aan van het zonder meer vervangen van ongemeten variabelen door indices om vervolgens te doen als betrof het een model met uitsluitend gemeten variabelen. In het onderhavige geval komt het bezwaar neer op het *ten onrechte niet korrigeren* van korrelatiekoëfficiënten voor invloeden van type a_1 meetfouten.

In de praktijk van het sociologisch en politikologisch onderzoek heeft men meestal te maken met modellen welke aanzienlijk komplexer zijn dan het hierboven behandelde model. De *strategie* welke hier voor de analyse van komplexe modellen wordt bepleit, kan worden verduidelijkt aan de hand van een model zoals weergegeven in figuur 6.5. Met opzet wordt hier de term strategie gebruikt, aangezien zich een dermate grote diversiteit van komplexe modellen kan voordoen dat een blauwdruk, in de zin van een handleiding, nauwelijks kan worden gegeven. Er zijn verschillende komplicerende elementen in bovengenoemd model aanwezig. Zo nemen we aan dat ξ_1, ξ_2, η_1 en η_2 staan voor *latente* variabelen. Voorts is er een methodenfaktor ξ_M aanwezig, welke zowel de indikatoren voor η_2 als die voor ξ_2 verstoort, is er sprake van een niet-rekursief model, en is er een samenhang tussen de meetfouten ε_{11}

en ε_{12}. Een andere moeilijkheid is dat in een dergelijk model gemakkelijk overidentifikatie optreedt, afhankelijk van het aantal indikatoren dat is geselekteerd. Tenslotte kan zich nog het probleem voordoen dat een latente variabele is opgebouwd uit meerdere dimensies. In het model hierboven kan η_1 als zodanig worden opgevat, met η_3, η_4 en η_5 als dimensies. Deze dimensies zijn gebaseerd op een verzameling indikatoren, welke ter vereenvoudiging zijn weggelaten en vervangen door de indices Y_3, Y_4 en Y_5. Hierop wordt aanstonds teruggekomen.



Figuur 6.5. Model ter verduidelijking van een procedure waarbij een LISREL-analyse wordt gekombineerd met de analyse van konseptmodellen.

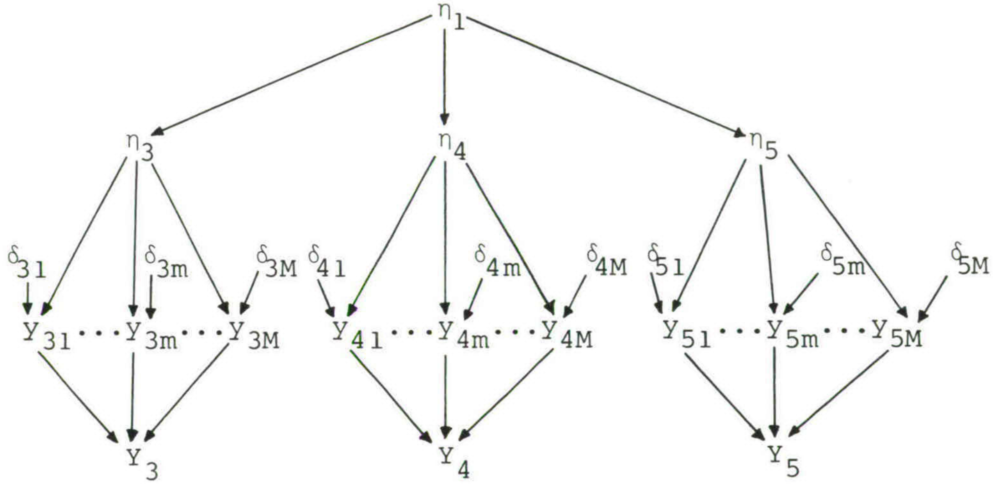
Gebruikmaking van regressie-analyse of pad-analyse voor het schatten van de modelparameters in een dergelijk model is uitgesloten wegens de aanwezigheid van niet-rekursiviteit, samenhang in de storingstermen (hier meetfouten), alsook een dreigende overidentifikatie. Ook instrumentele variabelen of "two stage least squares" kunnen hier geen uitkomst bieden. In een dergelijke situatie biedt Jöreskog's LISREL de beste oplossing, aangezien in deze procedure alle genoemde probleemfaktoren zijn toegestaan. Zoals in hoofdstuk 4 uitgewerkt, kunnen evenwel bezwaren worden ingebracht tegen LISREL, te weten het optreden van epistemische en strukturele bias alsook de mogelijkheid van het niet onderkennen van allerlei typen meetfouten. Vandaar dat wordt voorgesteld aan de LISREL-analyse vooraf te laten gaan een analyse van konseptmodellen.

Om te beginnen wordt het konseptmodel voor ξ_2 geanalyseerd. Exploratieve faktor-analyse op de indikatoren x_{2k} $k=1,\ldots,K$ levert de voor de berekening van de b_1 geldigheidskoëfficiënt $\hat{\theta}_{x2}$ benodigde faktorladingen op. Vervolgens wordt op basis van de genoemde indikatoren index X_2 samengesteld. Daarna komt X_2 in de plaats voor de indikatoren en wordt $\hat{\theta}_{x2}$ aangewezen als de epistemische parameter behorend bij de relatie tussen X_2 en ξ_2.

Vervolgens de analyse van de konseptmodellen voor η_2 en ξ_3. Gezien de werking van ξ_M kunnen deze konseptmodellen niet geheel gescheiden worden behandeld. De meest aangewezen werkwijze is hier één exploratieve faktor-analyse op de indikatoren y_{2p} $p=1,\ldots,P$ en x_{3l} $l=1,\ldots,L$. Op basis van de gevonden schattingen van faktor-ladingen kan vervolgens de verstorende werking van ξ_M worden geëlimineerd op een van de twee manieren zoals in de vorige paragraaf is voorgesteld onder het kopje "koppeling van konseptmodellen". Daarna kunnen op dezelfde wijze als hierboven de indices Y_2 en X_3 worden samengesteld en kunnen hun relaties met respektievelijk η_2 en ξ_3 worden voorzien van de als epistemische parameters fungerende b_1 geldigheidskoëfficiënten $\hat{\theta}_{y2}$ en $\hat{\theta}_{x3}$.

Tenslotte volgt nu de behandeling van het konseptmodel voor η_1. Zoals zojuist werd voorgesteld, wordt deze variabele opgevat als een meerdimensionele variabele met η_3, η_4 en η_5 als dimensies. Laten we voorts aannemen dat deze dimensies, evenals η_1 op te vatten als latente variabelen, zijn gebaseerd op de respektievelijke indikatoren y_{3m}, y_{4m} en y_{5m} m=1,...,M. (Met het oog op een vereenvoudigde notatie wordt uitgegaan van een voor elke dimensie gelijk aantal (M) indikatoren). Met behulp van deze indikatoren kunnen de indices Y_3, Y_4 en Y_5 worden samengesteld.

Een en ander kan worden voorgesteld als een konseptmodel zoals is weergegeven in figuur 6.6. Voor de analyse van modellen met een dergelijke formele struktuur ontwierp Jöreskog een speciale schattingsprocedure, welke bekend staat als "second order factor analysis". Bij deze procedure worden uit inputvariabelen faktoren getrokken, waarna uit deze faktoren zelf opnieuw faktoren worden gedestilleerd. Het is dus een faktor-analyse op een faktor-analyse.



Figuur 6.6. Konseptmodel voor de meerdimensionele latente variabele η_1.

Toepassing van deze procedure heeft evenwel de nadelen welke in hoofdstuk 4 tegen een dergelijke gang van zaken werden ingebracht. Daarom wordt voorgesteld het konseptmodel voor η_1

te verdelen in drie kleinere konseptmodellen, één voor elke dimensie. Op basis van deze modellen kunnen vervolgens de b_1-geldigheidskoëfficiënten θ_{y3}, θ_{y4} en θ_{y5} worden geschat en kunnen de indices Y_3, Y_4 en Y_5 worden samengesteld.

Na voltooiing van dit alles is voldoende voorarbeid verricht om met succes een LISREL-analyse uit te kunnen voeren. Deze verloopt als volgt. Als gefixeerde parameters worden ingevoerd $\hat{\theta}_{x2}$, $\hat{\theta}_{y2}$, $\hat{\theta}_{x3}$, $\hat{\theta}_{y3}$, $\hat{\theta}_{y4}$ en $\hat{\theta}_{y5}$ welke in het LISREL-programma als restrikties aan de schatting van modelparameters worden opgelegd. Vervolgens worden de indikatoren uit de konseptmodellen vervangen door de respektievelijke indices, waarna het LISREL-programma op de normale wijze van start kan gaan.

De kritiek waaruit de voorgestelde procedure is ontstaan, is in wezen een kritiek op gangbare wijzen van *modelspecifikatie*. Met name modellen welke zonder meer worden gespecificeerd met afzonderlijke indikatoren alsook modellen waarin indices samenvallen met de ongemeten variabelen welke zij beogen te meten, kunnen gezien de analyses in hoofdstuk 4 weinig vertrouwen inboezemen. Gelooft men wél in een model dat op een dergelijke wijze is gespecificeerd, dan kunnen de modelparameters zonder bezwaar op "full - information"-wijze worden geschat. In het geval dat er gegronde twijfels zijn ten aanzien van de juistheid van een dergelijke specifikatie levert een "full- information"-procedure een (overigens zuivere, konsistente en efficiënte) schatting van *niet - beoogde* parameters op.

Bij een analyse zoals werd voorgesteld, dient men zich af te vragen wat de semantische betekenis is van de variabelen in het model, of deze variabelen uit meerdere dimensies zijn opgebouwd en zo ja of deze dimensies theoretische relevantie hebben. In dit laatste geval dienen zij als aparte variabelen worden opgenomen in de hoofdtheorie, zoals hierboven bijvoorbeeld het geval was met η_3, η_4 en η_5. In het geval het gaat om ongeldige faktoren welke uitsluitend een rol spelen op het nivo van de hulptheorie, dienen korrelaties tussen indices voor hun werking te worden gekorrigeerd.

In gevallen waarin methodenfaktoren en ongeldige faktoren samenhangen, hetzij onderling hetzij met variabelen uit de hoofdtheorie, ontstaat meestal een dermate ingewikkeld relatiepatroon dat men veelal gedwongen zal zijn alsnog terug te vallen op de analyse van een LISREL-model waarin de indikatoren als *afzonderlijke* variabelen zijn opgenomen. Ditzelfde geldt wanneer als gevolg van de voorgestelde procedure het model ondergeïdentificeerd dreigt te geraken, hetgeen niet denkbeeldig is gezien de reduktie van het aantal variabelen in de hulptheorie die er het gevolg van is. Overigens kan de geschetste werkwijze behalve met LISREL óók worden gekombineerd met andere technieken zoals die van instrumentele variabelen en "two stage least squares".

Tot slot is een enkele waarschuwing op zijn plaats. De hierboven ontwikkelde procedures geven geen oplossing van het meetfoutenprobleem in elke willekeurige situatie die zich voordoet. Allereerst niet om een praktische reden. De geschetste werkwijze is in sterke mate afhankelijk van de beschikbaarheid van meerdere indikatoren per ongemeten variabele. Tenslotte is er nog een meer principiële reden. Deze wordt heel goed verwoord door Schoenberg: "Measurement error must eventually be eliminated at the source because the data-analysis methods for accounting for the presence of measurement error are ultimately only simplifications of the reality and consequently only approximately cover it up". Het tijdens de analyse-fase korrigeren voor meetfouten heeft dan ook altijd iets van een noodsprong, welke zoveel mogelijk vermeden moet worden.

SUMMARY

The analysis of causal models is one of the most widely-used methods of present-day social science research. In this study, causal models, which econometrists term structural equation models, are used to represent theories, with the result that causal analysis and problems of measurement and operationalisation become interwoven.

In this book this problem is examined by analysis of both the concept of measurement error and the semantic relationship between indicators and theoretical variables. Theoretical variables may be defined as those non-perceptible variables the concept of which is central to the analysis of measurement error; whilst measurement error is usually interpreted by social scientists as random interferences to the measurement process. But where theoretical variables relate to highly complex theoretical concepts, the interpretation of such concepts may lead to the development of nonrandom measurement errors in a latent variable. This interpretation process relates partly to the choice of indicators and partly to the auxiliary theory formulation. At the same time, the theoretical variables and their inter-relationship are set out in the principal theory. Thus, the purpose of analysing such models is to gain a better understanding of the principal theory from the perspective of relationships between the indicators as illustrated by use of the auxiliary theory.

Path analysis is one method which can be used in such cases. But there are many drawbacks in using this method for problems which involve the analysis of theoretical variables. These problems relate to overidentification, the absence of scores for theoretical variables, the correlation between independent variables and disturbances as well as

the correlation between the disturbances themselves. Path analysis does not provide a solution to any of these problems. Instead there is more advantage in using the instrumental variables or two-stage least-squares methods. Also useful are those estimating procedures which are based on the full-information maximum-likelihood principle, as, for example, LISREL.

But a problem related to the model-specification can develop in using these procedures. This problem arises if, as in most sociological and political science research, the indicators are treated as separate variables. This leads to conflicting structural and epistemological aspects of the model which is detrimental to the semantic qualities of theoretical variables. This effect follows from estimation being based on a matrix incorporating covariances between *all* indicators, with the result that epistemological parameters are partly determined by indicators of other theoretical variables in the model.

Thus the estimation of parameters for the principal theory can be adversely affected in this manner, notably through interference in the estimation of epistemological parameters. Another reason for this particular problem with the principal theory is that, unlike an index, an individual indicator provides only a limited representation of a theoretical variable. Thus if, in spite of this, estimation is based upon individual indicators, it will have a detrimental effect on the principal theory. Indices are therefor preferable to individual indicators as a basis for such estimation.

However, contrary to the belief of many social scientists, this does not mean that theoretical variables can therefore simply be replaced by indices. The epistemological parameters must be used during the analysis as a means of separating the index from the theoretical variable which it measures. Such parameters, which should comprise an aggregation of the epistemological parameters of the individual indicators, can be calculated through the application of factor-analysis to a concept-model. The resulting parameters may be likened to a reliability- or validity-coefficient and hence be very suitable for application as a fixed parameter in the development of models with theoretical variables. In other words, the numerical values of these coefficients can represent restrictions in estimating the free parameters.

Although a method of instrumental variables, like that of two-stage least-squares, permits the introduction of fixed parameters, Jöreskogs LISREL is considered the most useful procedure for this purpose. Its greatest advantage lies in the possibility of analysing nonrecursive models in addition to models containing errors in variables and errors in equations. Above all, this programme has the recognized advantages that are associated with a full-information maximum-likelihood estimation method. We may therefor anticipate that the application of LISREL, combined with an analysis of a concept model as elaborated in this book, will provide information with a validity of theoretical interpretation greater than any comparable procedures currently used for such purposes.

Bewijs dat λ_x in (5.14) gelijk is aan Ω van Heise en Bohrnstedt en vervolgens dat deze λ_x kan worden afgeleid uit de theoretische betrouwbaarheidskoëfficiënt (5.6.). Uitgaande van gestandaardiseerde inputvariabelen geldt in een orthogonale faktor-analyse: $\underset{\sim}{\Sigma}_{xx} = \underset{\sim}{\Lambda}\,\underset{\sim}{\Lambda}' - \underset{\sim}{I} + \underset{\sim}{H}^2$ waar $\underset{\sim}{H}^2$ staat voor een diagonale matrix met kommunaliteiten. In het geval van ongestandaardiseerde inputvariabelen moet deze vergelijking worden vervangen door

$$\underset{\sim}{\Sigma}_{xx} = \underset{\sim}{\Lambda}\,\underset{\sim}{\Lambda}' - \underset{\sim}{D}_\sigma^2 + \underset{\sim}{D}_\sigma \underset{\sim}{H}^2 \underset{\sim}{D}_\sigma$$

waar D_σ^2 staat voor een diagonale matrix met varianties van inputvariabelen. Substitutie van $\underset{\sim}{\Sigma}_{xx} + \underset{\sim}{D}_\sigma^2 - \underset{\sim}{D}_\sigma \underset{\sim}{H}^2 \underset{\sim}{D}_\sigma$ voor $\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}'$ in formule (5.14) geeft

$$\frac{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w} + \underset{\sim}{w}'\underset{\sim}{D}_\sigma \underset{\sim}{H}^2 \underset{\sim}{D}_\sigma \underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{D}_\sigma^2\underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}$$

$$= 1 - \frac{\underset{\sim}{w}'\underset{\sim}{D}_\sigma^2\underset{\sim}{w} - \underset{\sim}{w}'\underset{\sim}{D}_\sigma \underset{\sim}{H}^2 \underset{\sim}{D}_\sigma \underset{\sim}{w}}{\underset{\sim}{w}'\underset{\sim}{\Sigma}_{xx}\underset{\sim}{w}}$$

$$1- \frac{\sum_{k=1}^{K} w_k^2\ \sigma^2(x_k) - \sum_{k=1}^{K} w_k^2\ \sigma^2(x_k)\ h_k^2}{\sum_{k=1}^{K}\sum_{k=1}^{K} w_k w_l\ \sigma(x_k x_l)} = \Omega$$

Deze formule is afleidbaar uit die voor de theoretische betrouwbaarheidskoëfficiënt in (5.6). Wanneer in de teller hiervan $\sigma^2(\xi_k)$ wordt vervangen door de variantie van x_k vermenigvuldigd met de theoretische betrouwbaarheidskoëfficiënt voor indikator k $\left[\rho^2(x_k\xi)\right]$ dan volgt

$$\lambda_x^2 = \frac{\sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k)\,\rho^2(x_k\xi) + \sum\sum_{k\neq 1}^{K} w_k w_1 \sigma(x_k x_1)}{\sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) + \sum\sum_{k\neq 1}^{K} w_k w_1 \sigma(x_k x_1)}$$

$$= 1 - \frac{\sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) - \sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k)\,\rho^2(x_k\xi)}{\sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) + \sum\sum_{k\neq 1}^{K} w_k w_1 \sigma(x_k x_1)}$$

Aangezien geldt

$$\sum\sum_{k\neq 1}^{K} w_k w_1 \ \sigma(x_k x_1) = \sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) + \sum\sum_{k\neq 1}^{K} w_k w_1 \sigma(x_k x_1)$$

geeft dit na vervanging van theoretische itembetrouwbaarheden $\rho^2(x_k\xi)$ door kommunaliteiten h_k^2

$$\lambda_x^2 = 1 - \frac{\sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) - \sum_{k=1}^{K} w_k^2 \ \sigma^2(x_k) \ h_k^2}{\sum_{k=1}^{K} w_k w_1 \sigma(x_k x_1)} \qquad \square$$

NOTEN

Hoofdstuk 1.

1. Een kausaliteitsbegrip gebaseerd op een variabelentaal en het principe van verandering is niet het enige denkbare in politikologie en sociologie. Een andere mogelijkheid is dat slechts naar één eigenschap wordt gekeken om deze te volgen in de tijd. Men onderzoekt dan een evolutie zoals de ontwikkeling van een land van een diktatuur naar een demokratie. Men zoekt hier naar een verklaring van binnenuit, al dan niet ervan uitgaand dat het ene stadium van een ontwikkeling het volgende oproept. Nog een andere kausaliteitsopvatting is die van struktuur-oorzaak. Men onderzoekt hier bijvoorbeeld de vraag waarom zich in een maatschappij vervreemding kan voordoen. In deze gedachtengang moeten het bepaalde strukturele kenmerken zijn welke een maatschappij bevattelijk maken voor vervreemding. Met beide opvattingen van kausaliteit zal een empirisch onderzoeker zich niet licht bezig houden; zo ook ik in deze dissertatie, aangezien zij veel moeilijker toegankelijk zijn voor empirisch onderzoek dan kausaliteitsvragen welke passen binnen een variabelentaal.
2. Hierbij wordt onder andere gebruik gemaakt van komputerprogramma's voor regressie-analyse. Zo wordt in het SPSS-programma getracht op basis van F-toetsen op signifikantie van multiple korrelatie-koëfficienten, dan wel een toename daarin na invoering van een additionele regressor, te komen tot modelspecifikatie. Een nadeel van deze methode is het ad hoc karakter alsook de nadruk op regressie als voorspellingsmethode, dit laatste vanwege het grote belang dat wordt gehecht aan de proportie verklaarde variantie. Meer in overeenstemming met een kausale denkwijze is een modelspecifikatie gebaseerd op het kriterium van reproduktie van een korrelatiematrix op basis van modelparameters. Aanzetten hiervan zijn te vinden in de werken van Jöreskog.

Hoofdstuk 2.

1. Deze term is afkomstig uit de ekonometrie, waar men in plaats van met waarnemingen van verschillende eenheden op hetzelfde tijdstip meestal werkt met waarnemingen van dezelfde eenheden op verschillende tijdstippen. Vandaar deze term die wij liever onvertaald laten. Om dezelfde reden als de hier genoemde worden in de ekonometrie de

waarnemingen meestal aangeduid met subskript t (time) in plaats van de door ons gehanteerde n.

2. Het schattingsprobleem lijkt hier eveneens oplosbaar via toepassing van OLS op de gereduceerde vorm. Immers hier staan uitsluitend exogene variabelen in het rechterdeel van de vergelijking. Dit is inderdaad mogelijk, echter zulks uitsluitend in het uitzonderlijke geval dat het model exakt geïdentificeerd is. (Zie paragraaf 3 van dit hoofdstuk). Men spreekt in zo'n geval van "indirect least squares" (ILS).
3. Pad-analyse op ongestandaardiseerde skores is zonder meer uitvoerbaar. Men spreekt in dat geval van padregressie-analyse (zie Turner and Stevens 1959). De keuze hetzij voor gestandaardiseerde hetzij ongestandaardiseerde skores is niet zonder betekenis. Zie bijvoorbeeld Tukey 1954, Wright 1960 en Blalock 1966.
4. Wright zelf hanteerde hier steekproef-korrelaties. Zolang echter wordt gesproken op het niveau van modellen prefereer ik de notatie van populatie-korrelaties (Zie Wright 1954: 163)

Hoofdstuk 3.

1. Beide kategorieën van oplossingen zijn niet altijd even duidelijk gescheiden. Zo bewijst Christ dat 2SLS, welke methode zorgt voor een reduktie in het aantal schattingsvergelijkingen, in wezen ook kan worden gezien als een methode waarbij de alternatieve schattingen worden teruggebracht tot één unieke schatter. (Christ 1966: 446).
2. Een en ander geeft aanleiding tot de volgende overweging. Bij faktor-analyse wordt steeds de hele korrelatiematrix genomen als basis voor schatting. In termen van een IV-procedure: de vergelijkingen uit het faktormodel worden doorvermenigvuldigd met alle inputvariabelen behalve die welke als afhankelijke variabele staat in een vergelijking. Dit ongeacht de sterkte van de korrelatie van deze laatste met de overige inputvariabelen ofwel instrumentele variabelen. Vanuit IV-perspektief zou het beter zijn uitsluitend door te vermenigvuldigen met variabelen welke sterk korreleren. Dit zou neerkomen op hantering van een korrelatiematrix waaruit de over de hele linie zwak korrelerende variabelen zijn geëlimineerd. Nader onderzoek is hier gewenst.
3. Ik heb op dit punt lange tijd geaarzeld. Aanvankelijk stelde ik mij op het standpunt dat een keuze van instrumentele variabelen op basis van de sterkte van *een kausaal effekt* op de te vervangen variabelen het meest adekwaat is. Terwijl ik een dergelijk kriterium uitwerkte stootte ik op een publikatie van Fisher waarin eveneens een uitwerking van een dergelijk kausaal kriterium (Fisher 1965: 260-267). Volgens het door Fisher, en

onafhankelijk van deze eveneens door mij, uitgewerkt kausale kriterium wordt die IV geselekteerd welke een zo groot mogelijk (kausaal) effekt heeft op de te vervangen variabele. Bij nader inzien leek het mij dat die IV moet worden gekozen welke zich ten opzichte van zoveel mogelijk variabelen in het model exakt hetzelfde gedraagt (korreleert) als de te vervangen variabele. Het is niet zeker dat een kausaal kriterium dit kan garanderen. Vandaar dat ik alsnog koos voor het in de ekonometrie gangbare kriterium voor statistische samenhang. Nader onderzoek is hier evenwel gewenst.

Hoofdstuk 4

1. Om elk misverstand uit te sluiten; het gaat hier niet om externe indikatoren welke op *inhoudelijke* gronden een rol spelen in de betekenisverlening (type b_2 meetfouten) maar om invloeden welke optreden geheel en al als gevolg van de schattingsprocedure.
2. Het vooraf verenigen van indikatoren tot indices heeft nog een aantal additionele voordelen. Aldus ontstaat de mogelijkheid van standaardisering van meetprocedures en -instrumenten, het schatten per afzonderlijke vergelijking, het intuïtief beter begrijpen wat men aan het doen is en het verkrijgen van eenvoudiger modellen. Bovendien is aldus mogelijk een oplossing voor het probleem van overidentifikatie. Dit niet alleen vanwege het feit dat het aantal gemeten variabelen sterk wordt gereduceerd. Ook is het in principe mogelijk op basis van konseptmodellen de skores op latente variabelen te schatten waarna IV of 2SLS toepasbaar is in gevallen waarin dit nogmaal wordt verhinderd door ontbrekende skores. (Zie hoofdstuk 3).

Hoofdstuk 5.

1. In het onderstaande wordt voortgebouwd op een artikel van Heise en Bohrstedt waarin faktor-analyse wordt toegepast voor het bepalen van een betrouwbaarheids-, een geldigheids- en een ongeldigheidskoëfficiënt. (Heise and Bohrnstedt 1970). Met name worden deze koëfficiënten onderbouwd met een terminologie uit de meetfoutenproblematiek, (zie ook hoofdstuk 4 van deze studie), wordt de relatie met de klassieke meettheorie gelegd en worden de formules aanzienlijk eenvoudiger afgeleid dan het geval is bij de genoemde auteurs. Bovendien wordt aangetoond dat de afgeleide koëfficiënten op het nivo van indices kunnen worden opgevat als epistemische parameters.
2. Het is merkwaardig dat ook Jöreskog een betrouwbaarheidskoëfficiënt ontwikkelde welke (vrijwel) identiek is aan Ω zonder naar het toen reeds gepubliceerde werk van Heise en Bohrnstedt te verwijzen (Jöreskog 1971).

Een verschil met Ω is dat Jöreskog uitgaat van kongenerische items. In onze terminologie betekent dit dat Jöreskog moest uitgaan van afwezigheid van type a_2 en b_2 meetfouten, waardoor zijn koëfficiënt voor ons doel onbruikbaar is.

3. Het hier gemaakte onderscheid sluit goed aan bij de definities welke in de multitrait-multimethod theorie worden gegeven van de begrippen betrouwbaarheid en geldigheid. Betrouwbaarheid is de mate van overeenstemming tussen meetresultaten verkregen met maximaal *dezelfde* meetinstrumenten (indikatoren). Geldigheid is de mate van overeenstemming tussen meetresultaten verkregen met maximaal *verschillende* meetinstrumenten. Er is momenteel een trend merkbaar bij het analyseren van schalen waarbij dit onderscheid in typen indikatoren wordt genegeerd. Zo vindt men bijvoorbeeld in het SPSS pakket een toepassing van betrouwbaarheidskoëfficiënt α als kriterium voor het selekteren van schaal-items. Het is duidelijk dat dit niet juist is, aangezien schalen steeds zijn bedoeld voor de meting van latente variabelen. Het gaat niet aan om, op grond van geldigheidsoverwegingen geformuleerde, indikatoren (schaal-items) te selekteren op basis van een betrouwbaarheidskriterium.

Hoofdstuk 6.

1. Volgens Lawley en Maxwell is een faktormodel geïdentificeerd wanneer $s > o$ waar $s=\frac{1}{2}[(n-m)^2-(n+m)]$, n is het aantal inputvariabelen en m het aantal faktoren (Lawley and Maxwell 1971: 10). Uitgaande van een unifaktormodel (m=1) blijkt s gelijk aan nul te worden indien n=3.

LITERATUUR

Hoofdstuk 1.

Bertels, K. en Nauta D. (1969). Inleiding tot het modelbegrip. Bussum: De Haan.

Boudon R. (1971). Les mathématiques en sociologie. Parijs: Presses universitaire de France.

Braithwaite R.B. (1968). Scientific Explanation. A Study of the Function of Theory, Propability and Law in Science. Cambridge: University Press.

Bunge M. (1959). Causality. Havard University Press.

Carnap R. (1969). EinfUhrung in die Psylosophie der Naturwissenschaft. München: Nymphenburger Verlagshandlung.

Land K.C. (1971). Formal Theory. In Costner (1971). Sociological methodology. San Fransisco: Jossey Bass.

Margenau H. (1950). The Nature of Psysical Reality. A philosophy of Modern Psysics. New York: Mc Graw Hill Book Company Inc.

Melsen A.G.M. v. (1961). Causality and Determinism in Psysical Theory. In Nagel (1961). The Structure of Science. New York: Harcourt, Brace and World Inc.

Melsen A.G.M. v. (1962). Operationalisme en Causaliteit. Algemeen Nederlands Tijdschrift voor Wijsbegeerte en Psychologie, 54, 276286.

Nagel E. (1961). The Structure of Science. New York.

Pasquinelli A. (1967). Scientific Explanation and Causality. Quality and Quantity, 1, 32-52.

Welten W.P. (1961). Causaliteit in de Quantummechanica. Nijmegen: proefschrift.

Wippler R. (1969). Theory, Taxonomie en Model in de Sociologie. Mens en Maatschappij, 44, 271-286.

Zwart P.J. (1967). Causaliteit. Assen: Van Gorcum & comp. N.V.

Hoofdstuk 2.

Alwin D.F. and Hauser R.M. (1975) The Decomposition of Effects in Path Analysis. American Sociological Review, 40, 37-47.

Blalock H.M. (1966). Path Coefficients versus Regression Coefficients. THe American Journal of Sociology. dec.

Boudon R. (1971). A New Look at Correlation Analysis. In Blalock H.M. and Blalock A.B. (1971). Methodology in Social Research. London Mc Graw-Hill.

Christ C.F. (1966). Econometric models and methods. New York John Wiley and Sons Inc.

Duncan O.D. (1966). Path Analysis: Sociological Examples. American Journal of Sociology, 72, 1-16.

Duncan O.D. (1975). Introduction to Structural Equation Models. New York, Academic Press.
Duncan O.D., Haller A.O., Portes A. (1968). Peer influences on Aspirations: a Reinterpretation. American Journal of Sociology, 74, 119-137 en 75, 1042-1046.
Geer J.P. v.d. (1971). Introduction to Multivariate Analysis for the Social Sciences. San Fransisco: W.H. Freeman and Compagny.
Goldberger A.S. (1964). Econometric Theory. New York. John Wiley and Sons Inc.
Goldberger A.S. (1970). On Boudon's method of Linear Causal Analysis. American Sociological Review, 35, 97-101.
Goldberger A.S. (1972.b). Structural Equation Models in the Social Scienses. Econometrica, 40, 979-1001.
Goldberger A.S. (1973.a). Structural Equation Models: an Overview. In Goldberger A.S. and Duncan O.D. (1973). Structural Equation Models in the Social Scienses. New York. Seminar Press.
Harman H.H. (1970). Modern Factor Analysis. Chicago. The University of Chicago Press.
Hauser R.M. and Goldberger A.S. (1971). The treatment of Unobservable Variables in Path Analysis. In Costner H. (1971) Sociological Methodology. San Fransico. Jossey Bass.
Hays W.L. (1963). Statistics. New York. Rinehart and Winston Inc.
Heise D.R. (1975). Causal Analysis. New York. Wiley.
Johnston J. (1972). Econometric Methods. Tokyo: Mc Graw Hill.
Jöreskog K.G. (1976). Causal Models in the Social Sciences: the Need for Methodological Research. Research Report. Department of Statistics, University of Uppsala.
Land K.C. (1973). Identification, Parameter Estimation and Hypothesis testing in Recursive Sociological Models. In Goldberger A.S. and Duncan O.D. (1973). Structural Equation Models in the Social Sciences.
Linn R.L. and Werts C.E. (1969). Assumption in Making Causal Inferences from Part Correlations, Partial Correlations and Partial Regression Coefficients. Psychological Bulletin, 72, 307-310.
Mason R and Halter A.M. (1968). The Application of a System of Simultaneous Equations to the Innovation Diffusion Model. Social Forces, 47, 182-195.
Tukey J.M. (1954) Causation, Regression and Path Analysis. In Oscar Kempthorne (1954). Statistics and Mathematics in Biology. Iowa: Iowa State College Press.
Turner M.E. and Stevens C.D. (1959). The Regression Analysis of Causal Path. Biometrics, 15, 236-258.
Wold H.O. (1954). Causality and Econometrics. Econometrica, 22, 162-178.
Wold H.O. (1964). Econometric Model Building. Amsterdam: North Holland Publishing Company.

Wold H.O. (1969). Mergers of Economics and Philosophy of Science. Synthesis, 20, 427-482.
Wonnacott R.J. and Wonnacott T.H. (1970). Econometrics. New York. John Wiley and Sons Inc.

Hoofdstuk 3.

Alwin D.F. and Tessler R.C. (1974). Causal Models, Unobservable Variables and Experimental Data. American Journal of Sociology, 80, 58-86.
Blalock H.M. (1971). The Measurement Problem: a Gap between the Languages of Theory and Research. In Blalock H.M. and Blalock A.B. (1971). Methodology in Social Research. London: Mc Graw-Hill.
Blalock H.M. (1969). Multiple Indicators and the Causal Approach to Measurement Error. American Journal of Sociology, 75, 264-272.
Blalock H.M. (1971). Causal Models Involving Unmeasured Variables in Stimulus-Response Situations. In Blalock H.M. (1971). Causal Models in the Social Sciences. London: The MacMillan Press. Ltd.
Blalock H.M. (1975). The Confounding of Measured and Unmeasured Variables Sociological Methods and Research, 3, 255-383.
Burt R.S. (1973). Confirmatory Factor Analytic Structures and the Theory Construction Process. Sociological Methods and Research, 2, 131-190.
Christ C.F. (1966). Econometric Models and Methods. New York. John Wiley and Sons Inc.
Costner H.L. and Schoenberg R. (1973). Diagnosing Indicator Ills in Multiple Indicator Models. In Goldberger A.S. and Duncan O.D. (1973). Structural Equation Models in the Social Sciences. New York. Seminar Press.
Costner H.L. (1971). Theory, Deduction and Rules of Correspondence. In Blalock H.M. (1971). Causal Models in the Social Sciences. Chicago 1971.
Curtis R.F. and Jackson E.F. (1962) Multiple Indicators in Survey Research American Journal of Sociology, 68, 195-204.
Fisher F.M. (1965). The Choice of Instrumental Variables in the Estimation of Economy-wide Econometric Models. International Economic Review, 6, 245-274.
Goldberger A.S. (1971). Econometric and Psychometrics: a Survey of Communalities. Psychometrika, 36, 83-107.
Goldberger A.S. (1972.a). Maximum Likelyhood Estimation of Regressions Containing Unobservable Independence Variables. International Economic Review, 13, 1-15.
Goldberger A.S. (1973.b). Efficient Estimation in Overidentified Models: an Interpretative analysis. In Goldberger A.S. and Duncan O.D. (1973). Stuctural Equation Models in the Social Sciences.

Goldberger A.S. (1974). Unobservable Variables in Econometrics. In Zarembka P. (1974). Frontiers of Econometrics. New York: Academic Press.

Hauser R.M. and Goldberger A.S. (1971). The Treatment of Unobservable Variables in Path Analysis. In Costner H. (1971). Sociological Methodology. San Fransisco: Jossey Bass.

Jacobson A.L. and Lalu N.M. (1974). An empirical and Algebraic Analysis of Alternative Techniques for Measuring Unobserved Variables. In Blalock H.M. (1974). Measurement in the Social Sciences. Chicago.

Johnston J. (1972). Econometric Methods. Tokyo. Mc Graw Hill.

Jöreskog K.G. (1969.a). A General Approach to Confirmatory Maximum Likelyhood Factor Analysis. Psychometrika, 34, 183-202.

Jöreskog K.G. (1970). A general Method for Analysis of Covariance Structures. Biometrika, 57, 239-251.

Jöreskog K.G. (1973). A General Method for Estimating a Linear Structural Equation System. In Goldberger A.S. and Duncan O.D. (1973). Structural Equation Models in the Social Sciences. New York. Seminar Press.

Jöreskog K.G. (1976). Structural Equation Models in the Social Sciences: Specification, Estimation and Testing. Research Report. University of Uppsala.

Lawley D.N. and Maxwell A.E. (1971). Factor Analysis as a Statistical Method. London: Butterworths.

Long J.S. (1976). Estimation and Hypothesis-testing in Linear Models Containing Measurement Error. Sociological Methods and Research, 5, 157-206.

Malinvaud E. (1970). Statistical Methods and Econometrics. Amsterdam: NorthHolland Publishing Company.

Saris W.E. (1975). De Lisrel-procedure voor het Analyseren van lineaire Structurele Systemen met Ongemeten Variabelen. Sociale Wetenschappen, 17.

Siegel P.M. and Hodge R.W. (1971). A Causal Approach to the Study of Measurement Error. In Blalock H.M. and Blalock A.B. (1971). Methodology in Social Research, London. Mc Graw-Hill.

Werts C.E., Linn R.L. and Jöreskog K.G. (1971). Eastimating the Parameters of Path Models Involving Unmeasured Variables. In Blalock H.M. (1971). Causal Models in the Social Sciences. London: The MacMillan Press Ltd.

Werts C.E.. Jöreskog K.G. and Linn R.L. (1973). Identification and Estimation in Path Analysis with Unmeasured Variables. American Journal of Sociology, 78, 1469-1484.

Werts C.E., Linn R.L. and Jöreskog K.G. (1974). Quantifying Unmeasured Variables. In Blalock H.M. (1974). Measurement in the Social Sciences. Chicago.

Zellner A. (1970). Estimation of Regression Relationships Containing Unobservable Variables. International Economic Review, 11, 441-454.

Hoofdstuk 4.

Barman J.S. and Kenny D.A. (1976). Correlational Bias in Observer Ratings. Journal of Personality and Social Psychology, 34, 263-273.

Blalock H.M. (1965). Some Implications of Random Measurement Error for Causal Inferences. American Journal of Sociology, 71, 37-47.

Boesjes Hommes R.W. (1970). De Geldige Operationalisering van Begrippen. Meppel. Boom.

Bogue and Murphy. (1964). The Effect of Classification Errors upon Statistical Inference: a Case Analysis with Census Data. Demography, 1.

Bohrnstedt G.W. and Carter T.M. (1971). Robustness in Regression Analysis. In Costner H. (1971). Sociological Methodology. San Fransisco. Jossey Bass. Inc.

Burt R.S. (1973). Confirmatory Factor Analytic Structures and the Theory Construction Process. Sociological Methods and Research, 2, 131-190.

Campbell D.T. and Fiske D.W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin, 56, 81-105.

Galtung J. (1970). Theory and Methods of Social Research. London. George Allen and Union Ltd.

Thouless R.H. (1939). The Effects of Errors of Measurement on Correlation Coefficients. British Journal of Psychology, 29, 383-403.

Zellner R.A. and Carmines E.G. (1976) Factor Scaling, External Consistency and the Measurement of Theoretical Constructs. Political Methodology, 3, 215-252.

Hoofdstuk 5.

Alwin D.F. (1973). The Use of Factor Analysis in the Construction of Linear Composites in Social Research. Sociological Methods and Research, 2, 191-215.

Campbell D.T. and Fiske D.W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. Psychological Bulletin, 56, 81-105.

Cronbach L.J. (1951). Coefficient Alpha and the Internal Structure of Test. Psychometrika, 16, 297-334.

Heise D.R. and Bohrnstedt G.W. (1970). Validity, Invalidity and Reliability. In Borgatta and Bohrnstedt (1970). Sociological Methodology. London.

Long J.S. (1976). Estimation and Hypothesis-testing in Linear Models Containing Measurement Error. Sociological Methods and Research, 5, 157-206.

Lord F.M. and Novick M.R. (1974). Statistical Theories of Mental Test Scores. Addison-Wesley Publishing Company.

Mulder J., Pijper de W.M. en Saris W.E. (1977). Het Schatten van Factor Scores. Methoden en Data Nieuwsbrief, 2, 56-66.
Susmilch C.E. and Johnson W.T. (1975). Factor Scores for Constructing Linear Composites. Sociological Methods and Research, 4, 166-188.

hoofdstuk 6.

Jöreskog K.G. (1971). Statistical Analysis of Sets of Congeneric Tests. Psychometrika, 36, 109-133.
Lawley D.N. and Maxwell A.E. (1971). Factor Analysis as a Statistical Method. London. Butterworths.
Schoenberg R. (1973[2]). Strategies for Meaningful Comparison. In Costner (1972). Sociological Methodology 1972. London. Jossey-Bass Inc.
Wiley D.E. and Wiley J.A. (1974). A Note on Correlated Errors in Repeated Measurement. Sociological Methods and Research, 3, 173-183